# The Content-Driven Preprocessor of Images for MPEG-7 Descriptions

Jiann-Jone Chen, Cheng-Yi Liu† and Feng-Cheng Chang‡
Department of Electrical Engineering, National Taiwan University of Science & Technology
jjchen@mail.ntust.edu.tw

†Computer Commun. Labs., Industrial Technology Research Institute
bruceliu@itri.org.tw

‡Department of Electronic Engineering, National Chiao-Tung University
u8811833@cc.nctu.edu.tw

*Abstract*— **An image content-driven (CDP) preprocessor is proposed to activate the right MPEG-7 description tools for the recognized feature contents in one image. It determines automatically whether there are certain feature contents, such as color, texture or shape features, in one image and then performs processing to generate the corresponding descriptors. The CDP's most distinguished characteristic is that there are no redundant computations from the image content categorization down to the descriptor generation. Experiments show that the proposed CDP framework effectively categorizes images with accuracy up to 99%. We also proposed a practical content-based image retrieval (CBIR) system which integrate the CDP framework with a user-friendly MPEG-7 testbed. Simulations of CDP-based CBIR demonstrate that the CDP helps much in improving the subjective retrieval performance. This CBIR framework provide excellent flexibility such that it could be easily adapted to meet specific application requirements.**

## I. Introduction

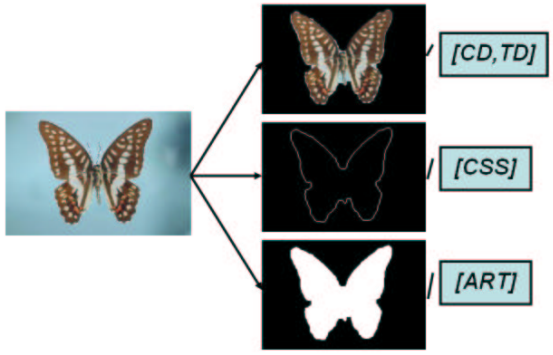
The multimedia content description interface, MPEG-7, provides normative descriptors for images, such as texture, color and shape et al, for effective similarity retrieval based on visual contents [1]. These descriptors represent visual contents with numerical feature values such that image resemblance could be measured quantitatively, i.e., by numerical distance values. The similarity retrieval for certain image database does perform well based on these descriptors. However, good retrieval performance could be targeted only when the key subjects of visual contents are precisely specified before description. For example, the shape descriptors should be imposed on the shape of meaningful objects in a trademark image instead of the blind description of the entire image. The color descriptors should be used to represent color features of the key foreground subject instead of the mixed features that involves the trivial background information. So that to design practical visual data retrieval system under the normative MPEG7, it's necessary to perform the pre-processing to let the right descriptors be utilized in the right content. For example, the information about object to be described within one image, such as contour or region as shown in Fig. 1 (a), must be available before performing further description process. In general, features (or descriptors) are extracted from images according to user's requirements, such as color histograms from color images and region shapes from trademark images et al. As there are more than one meaningful features could be extracted from one image, for clarity, we take the normative MPEG7 descriptors defined for images as the pre-processing targets. We propose to pre-categorize image feature genre [2] according to the acquisition and generating rules. As shown in Fig. 1 (b), the MPEG-7 description tools are activated whenever any specific content, i.e., shape, texture, color et al, are certified from the input image. It's reasonable that more than one set of descriptors could be extracted from one image.

Images are categorized by many [3]-[5] under various applications and specific requirements. For effective compression, images are categorized with neural network classifier according to which wavelet filter would perform better on a certain class of images [3]. A rough set [4] framework is proposed for CBIR applications, in which instead of using universal similarity measurements or manually selecting relevant features, it utilized the semantic content to generate rules which are then used

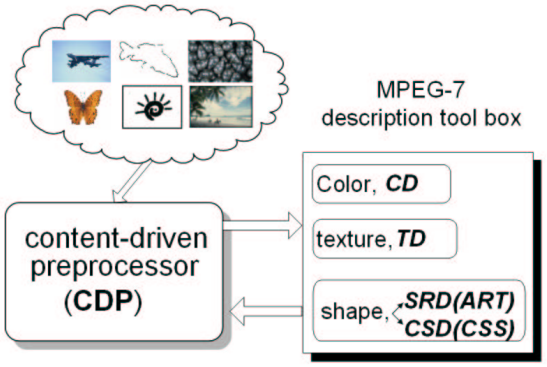
This work was partly supported by National Science Council of Taiwan under Grant No. NSC91-2218-E001-003, MOEA B32BXT5300 & LEE and MTI Center

to categorize images for advanced retrieval. Another one [5] hypothesizes that images with the same recurring patterns (or textures) are likely to belong to the same category. An $N \times M$ moving window is utilized to record the frequency profile, e.g., $N \times M-$ gram vectors, of one image for further similarity measurement. Unlike the categorization approaches above, we propose to exploit whatever feature contents in images that are certified by the MPEG-7 description. In practical implementation, we extract subjective MPEG7 descriptors and attach them to the image for comprehensive retrieval in any case. This CDP framework provides the efficient preprocessor for the user-friendly MPEG-7 CBIR testbed [6]. The integrated CBIR system could be easily adapted to satisfy user-specified application requirements.

The rest of this paper is organized as follows. The MPEG-7 normative descriptors are reviewed in Section II. The content-driven preprocessing is described in Section III. The simulation study is presented in Section IV. Section V concludes this paper.




1 (a)




1 (b)

Fig. 1. The preprocessing targets: (a) Recognized features (middle column) and associated descriptors (right column); (b) The content-driven preprocessor that activates corresponding MPEG-7 description tools for certified contents in one image.

## II. Image Description Tools

The MPEG-7 image description tools are reviewed in this section to help explaining the categorization mechanism in the CDP framework. The MPEG-7 normative visual descriptors for images are color, texture and shape which are briefly described below.

### A. Color Descriptor (CD)

The color description tools in MPEG-7 comprise color space, color quantization, dominant color, scalable color, color layout and color structure et al. These tools could be utilized for efficient retrieval only with the prerequisite that the subjective contents to be described are well-informed in advance. The color descriptors usually comprise color statistics such as histograms, i.e., $CD = [h(0), h(1), \cdots, h(n_C)]$, where $n_C$ is the number of bins specified to represent the range of color dynamics. In other words, the *CD* is just a feature vector represent the color properties of image contents.

### B. Shape Descriptor (SD)

The shape description tools for two-dimensional (2D) images comprise region-based and contour-based descriptors. For 2D binary images, the region shape descriptors are projections (or Angular Radial Transform (ART) coefficients) of shape contents to a unit disk with polar coordinates,

$$F_{nm} = \int_{\theta=0}^{2\pi} \int_{\rho=0}^{1} V^*_{nm}(\rho, \theta) \cdot f(\rho, \theta) \rho d\rho d\theta, \quad (1)$$

where $f(\rho, \theta)$ is the image signal function in polar coordinates and $V_{nm}(\rho, \theta)s$ are the ART basis functions with different order $m$ and repetition $n$ along radial and angular directions. The normalized Region-Shape Descriptors are $RSD = \{\frac{F_{nm}}{F_{00}}\}_{n=1,\cdots,N,m\leq n}$, where $N$ is the specified maximum number for order $n$.

The contour shape descriptors represent the counter image in the curvature scale space. It is basically the multi-resolution representation of contour curvature zeros. Let $\{(x(\sigma,t), y(\sigma,t)), t \in [0,1]\}$ denote the closed contour points, i.e., $(x(\sigma,0), y(\sigma,0)) = (x(\sigma,1), y(\sigma,1))$, with certain contour resolution $\sigma$. Larger $\sigma$s denote that more smoothness operations had been imposed on the contour shape. The curvature can be computed from the equation:

$$\kappa(\sigma,t) = \frac{x^{'}y^{''} - x^{''}y^{'}}{(x^{'2} + y^{'2})^{\frac{3}{2}}}(\sigma,t), \quad (2)$$

where $x^{'}$ and $x^{''}$ are the $1^{st}$ and $2^{nd}$ derivatives along $x$ directions and vice versa for $y$. The position of

zero-crossings can be determined by finding all pairs of consecutive indices $(t, t+1)$ for which $\kappa(\sigma, t) \cdot \kappa(\sigma, t) < 0$. The position of zero-crossings are recorded with normalized location $u$ (abscissa) and corresponding resolution $\sigma$ (ordinate). The contour shape descriptors are represented with $CSD = [c(0), c(1), \cdots, c(n_S)]$, where the coefficients are cyclically moved such that $c(0)$ is maximum and $n_S$ is the specified number of descriptors in $CSD$.

### C. Texture Descriptor (TD)

The texture description tools comprise homogeneous texture [7], texture browsing and edge histogram. For homogenous texture description, the frequency space is partitioned into 30 feature channels. The Radom transformed image is filtered with Gabor filters to get the frequency components in image for each channel. The energy $e_i$ and energy deviation $d_i$ of the frequency components constitute the texture descriptor:

$$TD = [f_{dc}, f_{sd}, e_1, e_2, \cdots, e_{30}, d_1, d_2, \cdots, d_{30}], \quad (3)$$

where $f_{dc}$ and $f_{sd}$ denote the brightness mean and standard deviation of the entire image pixels.

### D. Similarity Measurement by Descriptors

The similarity between two images in terms of one certain feature is measured by the numerical distance between their corresponding feature vectors. The distance computation is non-normative in MPEG-7 and we cab assign perception weighting for feature elements to evaluate subjective similarity. But the most critical operations in the indexing system would be the preprocessing. Above description tools are highly proficient in representing visual features for retrieval. However, they are activated under the knowledge of what visual features are certified. To bridge the gap of knowledge, we propose the **CDP** that it is designed to automatically activate the corresponding description tools according to image contents.

## III. Image Content-Driven Preprocessor

The two preprocessing targets are demonstrated in Fig. 1: (1) recognition of the image content features and associated descriptors; (2) activation of corresponding MPEG-7 description tools for certified contents in one image. In general, more than one set of descriptors could be extracted from one image, so the CDP is designed to extract whatever recognizable features from the image. As shown in Fig. 1 (a), the recognizable features of the butterfly image would be the region-shape, the contour-shape, the color and texture description of the butterfly.

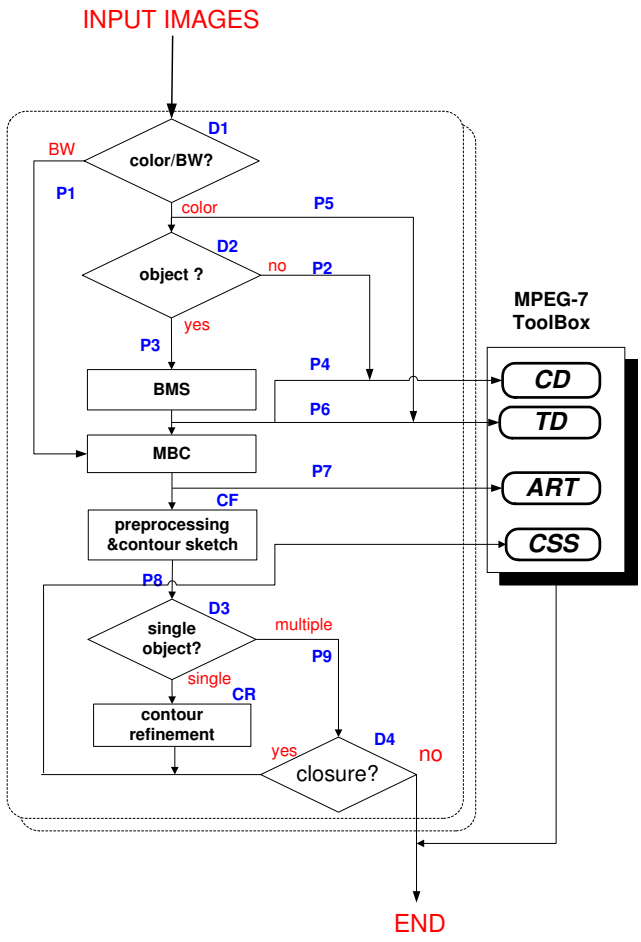



Fig. 2. The detailed control steps of CDP framework

These extracted features could further be evaluated by the confidence score for user query requirements [2].

For efficient computations in practical search engine, the CDP control is designed in a way that there're no redundant computations from deciding feature genre in one image to the descriptor generation, i.e., the computations for making the decision are also used in the description process. We describe the detailed CDP control with the aides of Fig. 2. The targets of the CDP are the normative MPEG-7 descriptors shown with bold round-cornered rectangles in the right portion. The other rectangles denote the well-developed procedures for specific requirements stated later. The diamonds are the decision functions to guide the categorization process. The path and decision number $P_i$ and $D_i$ are attached for easy understanding.

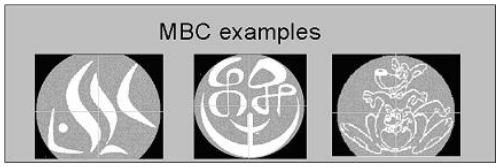
In general, only shape and contour information may exist in the black-and-white images, while color (or gray) image may additionally present texture and color information. So that the input images are first classified ($D_1$) into color (or gray) or black-and-white (BW).

The decision method is by first applying the two-layer moment preserving segmentation procedure [8] on the image and then compute the error distance $D_e$ between the segmented image and the original image with respect to the graylevels. Images with small $D_s$ would be determined to be black-and-white. If BW images are certified, the leftward path $P_1$ in Fig. 2 would lead to shape and/or contour description targets. If the input image is determined to be colored (or gray), either ($D_2$) there are discriminated foreground objects in image or the entire image content should be described altogether. In case of the later, color descriptors for the entire image are generated as the target ($P_2$). In case former, the background mesh segmentation (BMS,$P_3$) procedure is used to separate foreground from background. It then also activates the color description ($P_4$) for the segmented foregrounds. In addition to the texture descriptors for the entire image ($P_5$), the ones for foreground (if any) should also be extracted independently ($P_6$). Note that before texture description, it would determine whether there are salient homogeneous textures. The entropy of the TD, $H_{TD}$ in eq. 3 are computed to make the decision:

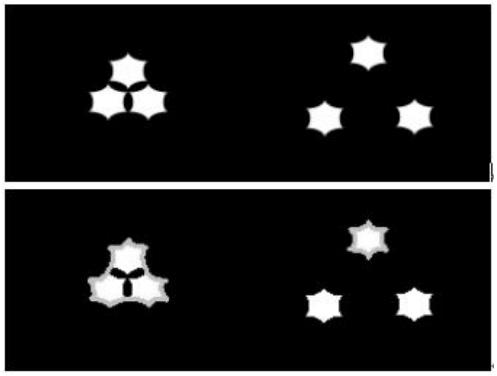
$$H_{TD} = -\sum_{i=1}^{30} p_i log p_i, \tag{4}$$

where $p_i = \frac{e_i}{\sum_{i=1}^{30} e_i}$. If the entropy is small enough ($\leq T_e$), i.e., there's salient homogeneous texture in image, then it activates the texture description.

Along path $P_1$ or downward path $P_3$, there resides region shape (ART) and contour shape (CSS) description tools for possible functioning. The decisions above, $D_1$ and $D_2$, conclude that there are shape contents in the image. Excluding the ones in BW images along path $P_1$, the foreground object downward $P_3$, after removing the image background, would also render shape and/or contour information. The MBC procedure is designed to find the minimum bounding circle of shape contents which would provide a normalized basis for the following shape descriptions. The shape descriptors ($P_7$) are computed from the region-shape image $f(\rho, \theta)$ according to eq. 1. We now discuss the possible contour description $P_8$. In the contour refining procedure CF, the image is processed with morphological operations for noise removal and for sketching the rough contour $C_O$ of region contents in which the size of structure elements is normalized based on diameter of the MBC. In decision $D_3$, if only one connected component (single) were detected, the sketched contour $C_O$ would be the only contour. Before feeding $C_O$ into the contour description, a refinement procedure CR would testify the royalty of $C_O$ to the original one shape content. A refined contour $C_F$ would be generated if needed. The closure test $D_4$ along $P_9$ is designed to test whether multiple regions are nested. If they were nested, then $C_O$ would be the ready contour for description.




3 (a) The MBCs



3 (b) The Closure Testing

Fig. 3. The well-defined procedures MBC and closure test in the CDP framework (a) The minimum bounding circle of region-shape images; (b) The closure test for region-shapes: upper row images are the original region-shape images and lower row images demonstrate the closure testing procedure. The lower left region-shape image satisfy the closure condition while the lower right not.

The detailed control steps of the MBC and closure testing are beyond the scope discussed here and are omitted. Only the processing functionalities are demonstrated in Fig. 3. In Fig. 3 (a), the white color regions are the original shapes and the shaded circle (not guaranteed to be a complete circle due to image aspect ratio) is the MBC for that region-shape in the image. The closure testing is designed under the assumption that only one key object can be defined in one image. In Fig. 3 (b), we can find a complete contour shape for the left image and cannot for the right one because these region-shapes are far from each other. In most cases, not to define a contour shape for the right region-shape image is reasonable. The left image thus pass the closure test while the right one not.

Note that $D_1$ is designed in a way that the generated testing image (moment preserved binary image) would be the shape image to be described if image is certified to be BW. The CF procedure is also designed in this way that the refined contour $C_F$, in addition to being used for testing, would be the final contour if certified. The procedures and decisions in the CDP are organized

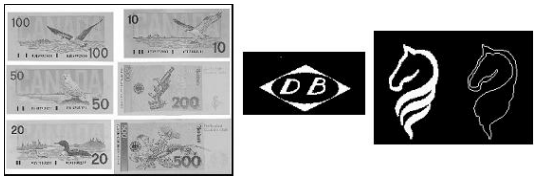


Fig. 4. The sample images that are misclassified

with the same target, i.e., no redundant computations.

## IV. Simulation Study

### A. Categorization Performance

Preliminary experiments on the categorization performance is demonstrate in Table 1. For JPEG-compressed BW images, the decision $D_1$ make 100% correct classifications. For (color) images with simple background, the decisions $D_1 \cdot D_2$ make up to 98% correct classifications. Note that the misclassified samples are with nearly indiscriminating background in the boundaries (e.g., the left image in Fig. 4) which could also be considered as no background. The contour images are almost classified correctly ($D_3 \cdot CR$) except one whose contour cannot match the conceptual horse head (the right image in Fig. 4. For sketching the contour of region contents, the samples are from the output of BMS and, in general, rough silhouettes exist in these samples. The missed samples are the incomplete region images resulting from preprocessing such that some thin lines in it would disappear (the middle image in Fig. 4). The precisions (or miss rate) listed in Table 1 are computed from the categorized samples in which the recognized descriptors could be clearly justified. Images which cannot be clearly categorized from human perception, such as texture images or images with obscure background, are not included in the image database.

| decisions | contents | certified | samples | correct | miss rate |
|---|---|---|---|---|---|
| $D_1$ | BW images | BW images | 26,700 | 26,700 | 0 % |
| $D_1 \cdot D_2$ | object | object | 864 | 854 | 1.16 % |
| $D_3 \cdot CR$ | contour | contour | 800 | 799 | 0.13 % |
| $D_3 \cdot D_4 / D_3 \cdot CR$ | object | contour | 599 | 548 | 8.35 % |

**Table 1.** The performance evaluation of the CDP.

### B. Application System Framework

In a typical content-based searching application, the CDP engine plays the feature extraction role. To construct a full-fledged CDP based application, we need sophisticated matching methods instead of plain low level feature combinations. The MPEG-7 testbed [6] provides a user-friendly environment to investigate and build content-based applications. The integration of the CDP engine and the MPEG-7 testbed is shown in Fig. 5. This integrated system could be easily adapted to fulfill various practical CBIR application requirements.

We believe the combination of the two efforts would improve the CDP performance and produce a practical CBIR system.

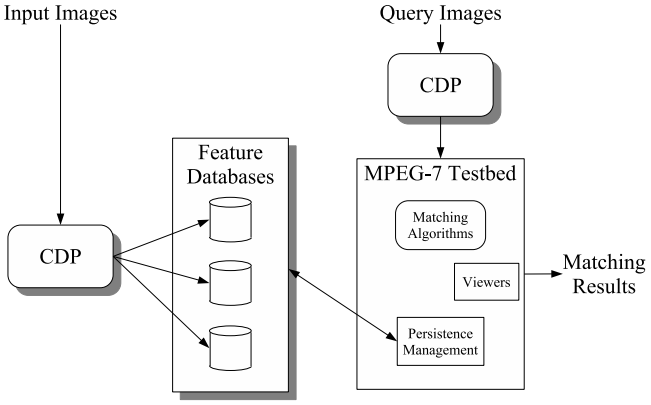



Fig. 5. Integration of CDP and Testbed

### C. Subjective Retrieval Performance

The subjective performance of the CDP-bassed CBIR is demonstrated in Fig. 6. In Fig. 6(a), if the descriptors are extracted from the entire image without categorization, it would lead to biased retrieval due to the dominated background color. The CDP-based CBIR eliminates the unnecessary backgrounds in the preprocessing and activates the corresponding description for the image content. In Fig. 6 (b), the two images in the left column are the query (top) and image features recognized by the CDP (below). As shown, the object colors dominate the retrieval such that some images with similar objects are retrieved. Results in Fig. 6 (c) are more accurate from subjective evaluation since right descriptors are used for retrieval. The other CDP-based CBIR results also demonstrate much improvement in subjective performance.

## V. Conclusion

An image content-driven preprocessor for MPEG-7 visual description has been proposed. It activates whatever description tools for the recognized features in images. The most distinguished characteristics of the CDP is that there's no redundant computations from the decisions down to the feature description. The categorization accuracy is up to 98%. In addition, the devised CDP is quite stable such that decision functions in the framework

could be improved or updated without any modification on the entire framework. For evaluating the efficiency of the CDP in practical CBIR applications, we integrate the CDP with a MPEG-7 testbed. Simulations demonstrate that the CBIR retrieval based on the CDP framework achieve good subjective performance. With the proposed CDP framework and the testbed, the MPEG-7 related retrieval capabilities could be largely improved.

## REFERENCES


[1] L. Cieplinski et al, "MPEG-7 visual part of eXperimentation model version 11.1," ISO/IEC JTC1/SC29/WG11 MPEG01/M7691, Dec. 2001.

[2] Cees G.M. Snoek and Marcel Worring, "A review on multi-model indexing," IEEE International Conference on Multimedia and Expo. ICME 2002.

[3] S. Saha et al, "Image categorization and coding using neural networks and adaptive wavelet filters," *IEEE Int. Conf. Indust. Tech. (ICIT)*, vol. 2, pp. 438-443, Jan. 2000.

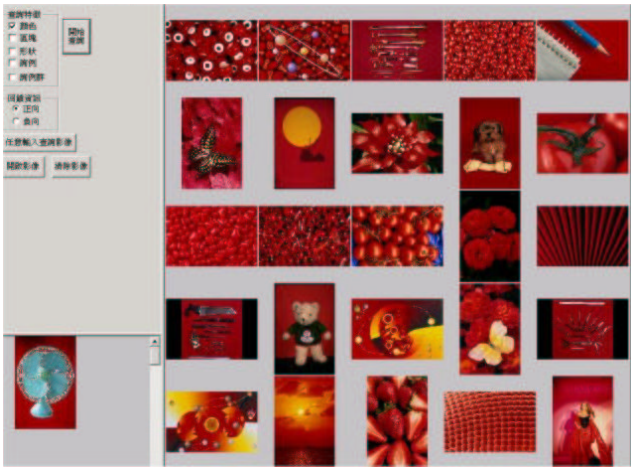
[4] V. K. Vutukuru et al, "A rough set framework for content based image classification and retrieval," in *Proc. Int. Conf. Multimedia Process. and Syst.,* Aug. 2000.

[5] A. Soffer, "Image categorization using $N \times M$ grams," *Proc. SPIE, Storage and Retrieval of Still Image & Video Databases V, pp. 121-132,* Feb. 1997.

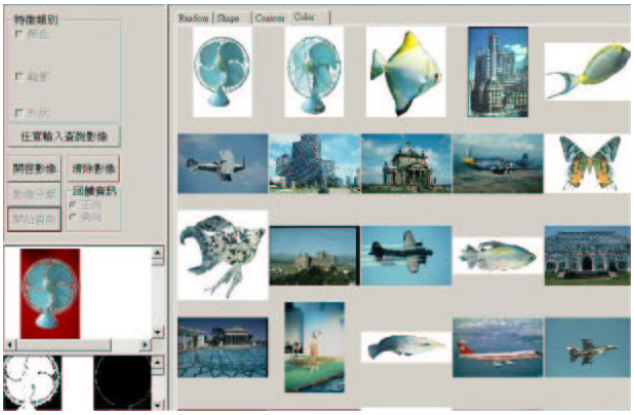
[6] F.-C. Chang, H.-M. Hang, and H.-C. Huang, "Research friendly MPEG-7 software testbed," in *Image and Video Communication and Processing Conf.*, Santa Clara, USA, Jan. 2003, pp. 890-901.

[7] Y. M. Ro et al, "MPEG-7 homogeneous texture descriptor," *ETRI Journal,* vol. 23, no. 2, pp. 41-51, June 2001.

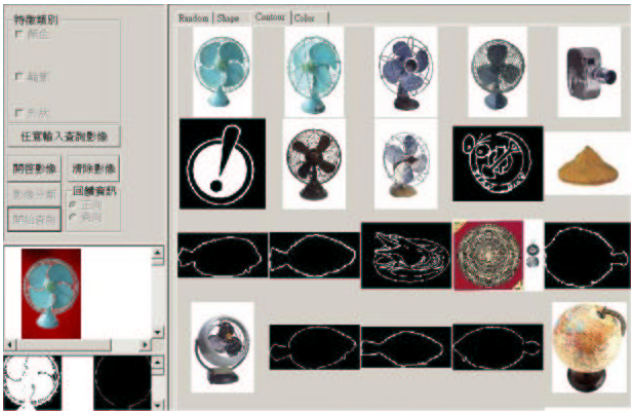
[8] W. S. Tsai, "Moment preserving thresholding, A new approach," in *Computer Vision, Graphics & Image Processing 29*, pp. 377-393, 1984

6 (a)



6 (b)



6 (c)

Fig. 6. The CDP-based CBIR: (a) biased retrieval due to dominated background color; (b) object-color dominated retrieval; (c) object-shape dominated retrieval.