퍼지적분을 이용한 내용기반 검색 사용자 의견 반영시스템

Relevance Feedback for Content Based Retrieval Using Fuzzy Integral

최 영 식*
Young Sik Choi

요 약

영상의 유사성에 대한 사용자의 주관적 인지를 학습하는 방법으로 relevance feedback 기술이 사용되며, 최근 들어 이에 대한 관심이 높아 지고 있다. 대부분의 relevance feedback기술은 영상 유사성을 측정하는데 사용되는 특징이 서로 독립적이라는 가정하고 있으나, 이러한 가정은 유사성 판단을 모델링 하는데 있어서 상당한 제약을 두는 것이다. 이 논문에서는, 퍼지 측정과 Choquet 적분을 이용하여, 유사성 판단에 대한 보다 나은 모델링 방법을 제안하고, 이를 이용한 relevance feedback 알고리즘을 제안한다. 실험결과를 통하여, 기존의 가중치 평균 방식에 의한 relevance feedback보다 제안된 방식이 우수함을 보인다.

Abstract

Relevance feedback is a technique to learn the user's subjective perception of similarity between images, and has recently gained attention in Content Based Image Retrieval. Most relevance feedback methods assume that the individual features that are used in similarity judgments do not interact with each other. However, this assumption severely limits the types of similarity judgments that can be modeled. In this paper, we explore a more sophisticated model for similarity judgments based on fuzzy measures and the Choquet Integral, and propose a suitable algorithm for relevance feedback. Experimental results show that the proposed method is preferable to traditional weighted- average techniques.

1. 서 론

With the recent advances in the multimedia technology and the advent of the World Wide Web, Content-Based Image Retrieval (CBIR) has become an actively researched area. Many visual feature representation techniques and systems have been built, e.g., [3][8][9][10]. One of the most important aspects of CBIR to be explored further isthe effective use of relevance feedback to improve the quality of retrieval. The goal of relevance feedback is to learn the user's subjective perception of similarity between images based on the user's interaction. The similarity paradigm used in the system needs to be sufficiently general and flexible so that it can be dynamically adapted to approximate a variety of similarity judgements encountered in practice. There are two basic types of relevance feedback: (1) weight updating [1][2][8][10], and (2) query point moving [11]. In the weight updating method, the weights (parameters) associated with the similarity measure are updated based on the feedback. In query point moving, the query is modified based on the feedback. In this paper, we consider the weight updating approach.

The problem of measuring the overall similarity between images with respect to multiple features can be considered as that of aggregation of evidence

* 종신회원 : 한국통신 멀티미디어연구소 영상미디어연구실장
choimail@kt.co.kr

obtained from different individual similarity metrics, one for each feature. The weighted average method is the commonly used aggregation method in CBIR systems due to its simplicity [1][2][10]. This paradigm, however, is not flexible enough to model queries such as: Retrieve all images that are similar to the given image with respect to any 3 features out of 5. In this case, the aggregated value should be high whenever any three or more individual metrics provide high similarity values. This cannot be modeled by a weighted average since we do not know which three features are applicable for any given image. Using equal weights for all 5 features does not give high enough values for the overall similarity. Moreover, the weighted average paradigm assumes that the features are independent of each other so that the overall similarity can be modeled as the sum of (weighted) individual similarities. However, in general features are not independent of each other and we need to consider interactions between features. For example, some of the features might be highly correlated, and therefore, the weight for the combined evidence from these features should be less than the sum of the weights for the individual features. Also, similarity with respect to a particular feature may be important to the user only if similarity with respect to some other feature is present or absent.

The Choquet Integral (CI) [12] has been proposed [4] as a similarity measure in order to overcome the above problems with the weighted average aggregation. (See also [5] for a related measure.) In this approach, a fuzzy measure is defined over the power set of a given set of features and CI is used to aggregate the individual similarities with respect to the fuzzy measure. Therefore, interactions between features can be represented and highly sophisticated queries can be modeled. However, relevance feedback is not discussed in [4]. In order to utilize the full potential of CI for CBIR, we need to estimate the fuzzy measure to be used for a given query based on the relevance feedback provided by the user. Moreover, the estimation process needs to be fast when we are dealing with a very large image database or with the Web [8]. In [7] an algorithm for relevance feedback based on the Sugeno measure is suggested. However, the Sugeno measure is highly restrictive in that it assumes that all pairs of features interact the same way. In this paper, we explore the use of the CI as a similarity measure, and propose a practical algorithm to dynamically estimate the fuzzy measure (in its most general form) based on the feedback provided by the user.

2. Fuzzy Measure and Choquet Integral

2.1 Fuzzy Measure

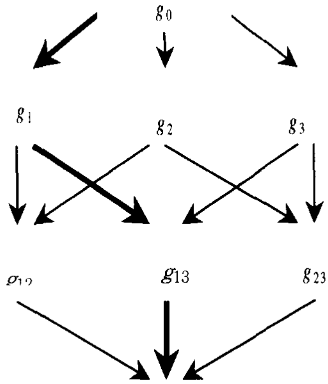
Let $X = \{x_1, x_2, x_n\}$ denote a finite universal set and let $P(X)$ denote the power set of X. A function $g : P(X) \rightarrow [0, 1]$ that satisfies the following properties is called a fuzzy measure:

(i) $g(\varnothing) = 0$, $g(X) = 1$, and

(ii) If $A, B \subset X$ and $A \subset B$, then $g(A) \subset g(B)$

Note that a fuzzy measure $g(\cdot)$ requires 2^n coefficients, i.e. the $g(\cdot)$ values of the 2^n subsets of X. The value of $g(A)$ denotes the worth or importance of subset A. If a fuzzy measure g is additive, then $g(\{x_1, x_2\}) = g(\{x_1\}) + g(\{x_2\})$, i.e., measure of the whole is the sum of the measures of the parts. If we allow interactions between x_i's, then measure of whole may be less than or greater than the sum of the measures of the parts.

The monotonicity requirement in (ii) above can be depicted in a lattice. Figure 1 shows the lattice for X



(Figure 1) Lattice of coefficients of a fuzzy measure

= $\{x_1, x_2, x_3\}$, where g_{ij} denotes $g(\{x_i, x_j\})$. The lattice has n+1 horizontal layers. The layer containing g_0=$g(\varnothing)$ is layer 0 and the layer containing g_X=$g(X)$ is layer n, where n is the cardinality of the universal set X. A path is defined as a sequence of g's, starting from g_0 and ending in g_X. Property (ii) requires that the measure is always increasing when we travel along a path. For a given node in layer l representing $g(A)$, its lower neighbors (upper neighbors) are defined to be the set of nodes in layer l - 1 (l + 1) that represent the measures of $g(B)$ where B is a subset (superset) of A. There are l lower neighbors and n - l upper neighbors for a node in layer l.

2.2 Choquet Integral

Let f : X [0, 1] be a function and let g be a fuzzy measure on X. The Choquet Integral (CI) of f with respect to g is defined by

$$C_g(f) = \int_X f \circ g = \sum_{i=1..n}[f(x_{(i)}) - f(x_{(i-1)})]g(A_{(i)}) \quad (1)$$

where $x_{(i)}$ denotes the i-th element of X when the elements are permuted such that $0 \le f(x_{(1)}) \le f(x_{(2)}) \le f(x_{(n)}) \le 1$, and $A_{(i)} = \{x_{(i)}, , x_{(n)}\}$, i = 1, ..., n, are a nested sequence of subsets of X. In (1), $f(x_{(0)})$ is defined to be 0. In the present application, $f(x_i)$ stands for the similarity between two images with respect to feature x_i. For a given f, the computation of the integral involves only one path in the lattice structure. For example, the path corresponding to the case when $f(x_2) \le f(x_3) \le f(x_1)$ is shown in Figure 1 in bold. CI can also be computed by the following equivalent formulation:

$$C_g(f) = \int_X f \circ g = \sum_{i=1..n} f(x_{(i)}) [g(B_{(i)}) - g(B_{(i-1)})] \quad (2)$$

where x_i's are now arranged so that $1 \ge f(x_{(1)}) \ge f(x_{(2)}) \ge f(x_{(n)}) \ge 0$, $B_{(i)} = \{x_{(1)}, \ldots, x_{(i)}\}$, and $B_{(0)} = \varnothing$. In this formulation, $f(x_{(i)})$ is multiplied by the increase in worth or importance, i.e., by $g(B_{(i)})$ $g(B_{(i-1)})$, when $x_{(i)}$ is added to the pool $B_{(i1)} = \{x_{(1)}, , x_{(i-1)}\}$. If $x_{(i)}$ brings little new information, then the measure should be chosen such that $g(B_{(i)}) - g(B_{(i-1)}) \ge 0$.

The Choquet integral is reduced to a weighted average when the fuzzy measure g is additive. That is, if $g(\{x_{(1)}, \ldots, x_{(i)}\}) = g(\{x_{(1)}\}) + \ldots + g(\{x_{(i)}\})$, then $g(B_{(i)}) - g(B_{(i-1)}) = g(\{x_{(i)}\})$. Therefore, (2) becomes

$$C_g(f) = \sum_{i=1..n} f(x_{(i)}) \, g(\{x_{(i)}\}) = \sum_{i=1..n} f(x_i) \, g(\{x_i\}) \quad (3)$$

If $g(B)$ depends only on the cardinality of B, i.e., if $g(A) = g(B)$ whenever $|A| = |B|$, then $g(B_{(i)}) - g(B_{(i-1)}) = g(|B_{(i)}|) - g(|B_{(i-1)}|) = g(i) - g(i\text{-}1)$ will be independent of the specific sets $B_{(i)}$ and $B_{(i-1)}$, and can be written as w_i. Equation (2), therefore, can be reduced to the ordered weighted average (OWA) operator [6].

$$C_g(f) = \int_X f \circ g = \sum_{i=1..n} f(x_{(i)}) \, w_i \quad (4)$$

It is well-known that OWA can yield any order statistic (such as the median) with a proper choice of w_i, $i = 1, ..., n$. Also, it can simulate concepts such as at least k out of n or any k out of n. It can be easily shown [6] that

$$\min_i f(x_{(i)}) \leq C_g(f) \leq \max_i f(x_{(i)}) \quad (5)$$

3. Relevance Feedback

3.1 Aggregation of Similarity Measures

Let $X = \{x_1, ..., x_n\}$ be a set of n features where x_i represents i-th feature, for example, color histogram or wavelet coefficients. Let $S = \{s_1, ..., s_n\}$ be a set of similarity measures between two images with respect to features $X = \{x_1, ..., x_n\}$, where $s_k(\cdot)$ represents a similarity measure with respect to feature x_k. For example, X can be {color, texture, shape} and s_1, s_2 and s_3 can be a set of similarity measures with respect to color, texture and shape. Let x_{qk} and x_{jk} denote k-th feature of the query image q and of image j in the database, respectively. Then, we can define the overall similarity between the images as the Choquet integral that combines the similarities using an appropriate measure, i.e.,

$$C_g(S, q, j) = \sum_{k=1..n}[s_{(k)}(x_{q(k)}, x_{j(k)}) - s_{(k-1)}(x_{q(k-1)}, x_{j(k-1)})]g(A_{(k)}) \quad (6)$$

where $0 \leq s_{(1)}(x_{q(1)}, x_{j(1)}) \leq ... \leq s_{(n)}(x_{q(n)}, x_{j(n)}) \leq 1$. As described in Section 2, in this paradigm, we can model highly sophisticated similarity judgements rather than simple weighted averages.

3.2 Learning feature relevance

In (6), $g(A_{(k)})$ can be considered as the importance or relevance of the feature subset $A_{(k)} = \{x_{(1)}, ..., x_{(k)}\}$. Therefore, the problem of learning the similarity measure based on relevance feedback is reduced to updating an initial default measure $g(\cdot)$ so that the measure is biased toward more relevant images. Let R^+ be the set of relevant examples and R^- the set of irrelevant examples indicated by the user. Using R^+ and R^-, we need to update $g(\cdot)$ so that the images in R^+ are ranked high and images in R^- are ranked low. We achieve this by updating the fuzzy measure coefficients $g(A_{(k)})$ that participate in (6) for a given query image q so that for samples i in R^+, $C_g(S, q, i)$ is increased and for samples j in R^-, $C_g(S, q, j)$ is decreased. Specifically, we minimize:

$$J(g) = \sum_{i\in R+} [C_g(S, q, i) - s_{(n)}(x_{q(n)}, x_{i(n)})]^2 + \sum_{j\in R-} [C_g(S, q, j) - s_{(1)}(x_{q(1)}, x_{j(1)})]^2 \quad (7)$$

where $s_{(n)}(x_{q(n)}, x_{i(n)})$ is the maximum among the individual similarity measures $\{s_1(\cdot), ..., s_n(\cdot)\}$ between the query image q and the database image i, and $s_{(1)}(x_{q(1)}, x_{j(1)})$ is the minimum among the individual similarity measures (see also (5)).

We cannot blindly apply gradient descent while minimizing J in (7) with respect to $g(A_{(k)})$ because of the monotonicity constraints on the fuzzy measure coefficients (see Section 2). In addition to this difficulty, for a given image i, the computation of the integral involves only one path in the lattice as shown in Figure 1, and in general the number of images that are involved in the feedback is small. Therefore, there will be many unmodified nodes in the lattice structure when the gradient descent method is applied. We need to deal with these unmodified nodes and propagate errors through the whole lattice structure for better generalization. We do this by modifying the algorithm proposed by Grabisch in [5], which was developed for pattern recognition problems. In pattern recognition, the objective function is different from (7), and learning (training) need not occur in real time. In relevance feedback, we need to train the fuzzy measure coefficients on the fly and retrieve images using trained coefficients as quickly as possible. The algorithm is as follows.

1. For a given query q and a set of the feedback examples R^+ and R^- given by a user, do the following.
2. Identify the path and compute the error for each feedback example p. Let $g(0), g(1), ..., g(n)$ denote the fuzzy measure coefficients involving the path. For example, in Figure 1, we have $g(0) = g_0$, $g(1) = g_1$, $g(2) = g_{13}$, $g(3) = g_{123}$. Using (7), the errors can be computed as

$$E = C_g(S, q, p) - s_{(n)}(x_{q(n)}, x_{p(n)}) \text{ if } p \in R^+$$
$$E = C_g(S, q, p) - s_{(1)}(x_{q(1)}, x_{p(1)}) \text{ if } p \in R^-$$

3. Update the fuzzy measure coefficients in the path using the gradient descent method as follows:

$$g(k) = g(k) + \varepsilon E[s_{(k)}(x_{q(n-k)}, x_{p(n-k)}) - s_{(k-1)}(x_{q(n-k-1)}, x_{p(n-k-1)})],\ k = 1,...,n.$$

where ε denotes the learning rate.
4. Enforce the monotonicity constraints for the updated fuzzy measure coefficients in Step 3 as follows: When $E > 0$, starting from $g(1)$ to $g(n\text{-}1)$, if $g(k) < \max LN(k)$, then set $g(k) = \max LN(k)$, where $\max LN(k)$ is the maximum value of the fuzzy measures in lower neighbors of $g(k)$. When $E < 0$, starting from $g(n\text{-}1)$ to $g(1)$, if $g(k) > \min UN(k)$, then set $g(k) = \min UN(k)$, where $\min UN(k)$ is the minimum value of the fuzzy measures in upper neighbors of $g(k)$. See Section 2.1 for the definitions of lower and upper neighbors.
5. Update the unmodified fuzzy measure coefficients $g(k)$'s in the previous steps, starting from the nodes in the lowest layer, as follows:

$$g(k) = (\text{avg}LN(k) + \text{avg}UN(k))/2.0,$$

where $\text{avg}LN(k)$ and $\text{avg}UN(k)$ are the average values of the fuzzy measure coefficients of lower neighbors and upper neighbors of node k, repectively. After updating $g(k)$, enforce the monotonicity constraint as in Step 4.
6. Repeat step 2 through 5 until $g(\cdot)$ converges. The learning rate decreases in each iteration so that it is guaranteed to converge.

Note that in step 5, unmodified nodes are defined to be the nodes that have not been modified since the beginning of the relevance feedback session for a given image q. For the initial query from the user, the individual features are considered to be equally important and independent, and $g(\{x_i\})$ are initialized to $1/n$ for i=1,...,n. The other fuzzy measures in the lattice are computed assuming that $g(\cdot)$ is additive (See section 2). During the entire relevance feedback session for a given query, we do not re-initialize the $g(\cdot)$ values, but continually update them.

Since $g(\cdot)$ involves 2^n coefficients, learning can be slow when n is extremely large. One solution is to combine the individual similarity metrics at a lower level using a simple method (e.g. weighted average), and apply CI to the aggregated metrics at a higher level. We now illustrate this procedure.

4. Experimental Results

We tested the proposed approach over two different image data sets: (i) the Vistex data set where the ground truth is known and (ii) a collection of

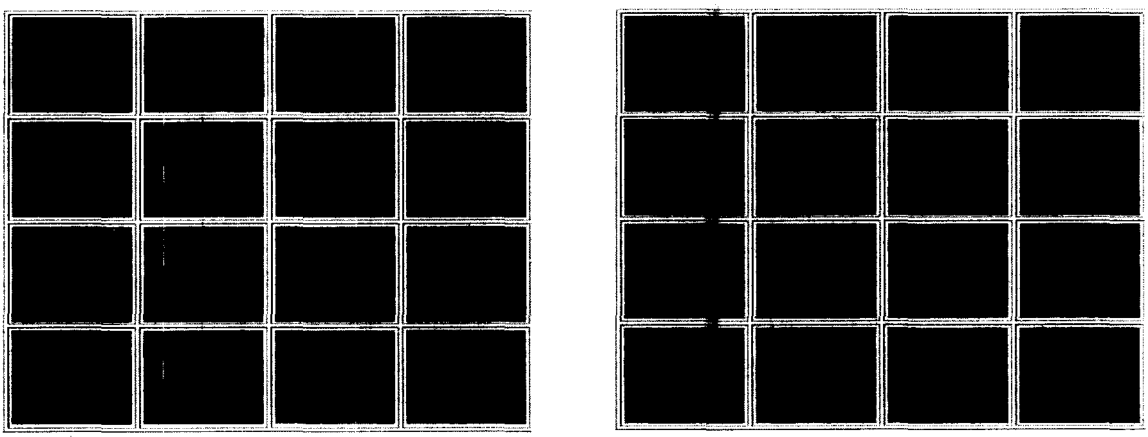
1600 images obtained from various sources, such as stock photo agents and the Web. There were three types of features used for this experiment, i.e., 196 features extracted from the color histogram, 256 features extracted from the color and texture layout, and 120 texture features based on the wavelet transform. The description of these features is outside the scope of this paper. We combine the similarities based on the color histogram features into one measure, and this measure is denoted by C. Similarly, the combined measure from the layout features is denoted by L and the combined measure from the texture features is denoted by T. The fuzzy measure is defined over for the set $\{C, L, T\}$. In the experiments, we set the learning rate to 1 in Step 3 of the algorithm, and the maximum number of times Steps 2-5 are repeated in the algorithm to 10.

In the case of Vistex, the original 210 texture images of size 512×512 were further divided into 16 sub-images of size 128×128 to create a database of 3360 images. Therefore, we assume that the 15 siblings are relevant for a given 128×128 query image. Table 1 shows a comparison of the performance of the CI approach with the weighted average (WA) method adapted from [1] on the Vistex data set. Since the precision for the initial retrieval is already higher than 90% in most cases, we only show the results for the cases with precision less than 70%. The results show that CI does better than WA.

(Table 1) A Comparison of the Retrieval Precision of CI and WA methods on the Vistex data set

Method	No. times relevance feedback is sought from user				
	0	1	2	3	4
CI	52.96	71.05	74.67	76.64	78.9
WA	52.96	54.60	58.22	60.19	63.20

We illustrate the CI method with an example. Figure 2(a) shows 16 retrieved images before relevance feedback. Figure 2(b) shows the improved retrieval after 2 feedback iterations. Table 2 shows the corresponding fuzzy measure, where the first row corresponds to layer 1 in Figure 1, and the second row to layer 2. The initial (default) coefficients are shown in parentheses in each case. It can be seen that the



(a) (b)

(Figure 2) Relevance feedback Example: (a) Retrieval result without relevance feedback and (b) Improved result after two feedback iterations

(Table 2) Fuzzy measure after 2 feedback iterations for {C=color, L=layout, T=texture}, and default values

$g(\{C\})$ = 0.369 (0.333)	$g(\{L\})$ = 0.374 (0.333)	$g(\{T\})$ = 0.788 (0.333)
$g(\{C, L\})$ = 0.686 (0.667)	$g(\{C, T\})$ = 0.789 (0.667)	$g(\{L, T\})$ = 0.812 (0.667)

paths $g(\{T\}) \rightarrow g(\{C, T\}) \rightarrow g(\{C, T, L\})$ and $g(\{T\}) \rightarrow g(\{L, T\}) \rightarrow g(\{C, T, L\})$ are significantly changed. In these paths, the weight for similarity based on feature T is 0.788. When feature L is added to T, the weight increases by 0.024, and when feature C is added, it increases by 0.001. This indicates that similarity with respect to feature C is not important when similarity with respect to feature T is strong. When the third feature is added to the above two cases, the weight increases by 0.188 and 0.211 respectively (see equation (2)). The interpretation is that feature T is very important and two other features are of moderate importance in the similarity judgement of the user. Figure 2(b) illustrates this behavior. The weighted average method does not work as well, as can be seen from Table 1.

Since the ground truth for the second 1600-image database is not available, we performed subjective evaluations of CI and WA methods for this database. The overall results show that CI is very effective. A demo of these two methods is available on the Web (http://210.105.44.171/choi).

5. Conclusions

We proposed a relevance feedback mechanism for CBIR based on the Choquet integral. In this approach, a fuzzy measure is defined over the set of features and the Choquet Integral is used to evaluate the similarity between images. By taking into account the interdependencies between the features, CI can model highly sophisticated similarity judgements. We presented an algorithm to estimate the fuzzy measure on the fly based on the user's feedback. Results on the Vistex data set and a collection of 1600 real images show that the proposed method performs better than the traditional weighted average method.

References

[1] Yong Rui, *et al*, Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval, *IEEE Trans. Circuits and Video Technology*, 8(5): 644-655, 1995.

[2] B. Bhanu, J. Peng, and S. Quing, Learning Feature Relevance and Similarity Metrics in Image Databases. *IEEE Workshop on Content-Based Access of Image and Video Libraries*, Santa Barbara, CA. USA 1998, pp. 14~18.

[3] W. Y. Ma and B. S. Manjunath, Netra: A Toolbox for navigating large image databases, *Proceedings of IEEE Int. Conference on Image Processing*, 1997.

[4] S. Santini and R. Jain, Similarity Measures, *IEEE Transactions on PAMI*, Vol. 21(9): 1999, pp. 871~883.

[5] M. Grabisch, A New Algorithm for Identifying Fuzzy Measures and Its Application to Pattern Recognition, Proc. *FUZZIEEE*, Yokohama Japan, 1995.

[6] M. Grabisch, On Equivalence Classes of Fuzzy Connectives: The Case of Fuzzy Integrals, *IEEE Transactions on Fuzzy Systems*, 8(1): 96-109, 1995.

[7] H. Frigui, Adaptive Image Retrieval Using the Fuzzy Integral, Proc. *NAFIPS*, New York, 1999, pp. 575~578.

[8] S. Scalaroff, L. Taycher, and M. La Cascia,

Imagerover: A Content Based Image Browser for the World Wide Web, *IEEE Workshop on Content Based Access of Image and Video Libraries,* June 1997.

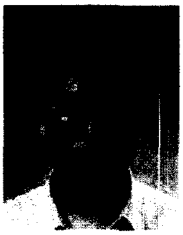
[9] W. Niblack, R. Barber, and et al., The QBIC project: Querying images by content using color, texture and shape, *Proc. SPIE Storage .and Retrieval for Image and Video Database*, Feb 1994.

[10] J. R. Bach, C. Fuller, and et al., The Virage image search engine: An open framework for image management, *Proceeding of SPIE Storage and Retrieval for Image and Video Database*, 1995.

[11] C. Meilhac and C. Nastar, Relevance feedback and category search in image databases, *Proc. IEEE Int. Conf. Multimedia Comp. & Syst.*, Florence, 1999, pp. 512~517.

[12] T. Murofushi and M. Sugeno, An interpretation of fuzzy measure and the Choquet integral as an integral with respect to a fuzzy measure, *Fuzzy Sets and Systems*, Vol. 29, 1989, pp. 201~227.

◐ 저 자 소 개 ◑



YoungSik Choi received the B.S. and M.S. degrees in electronics from Yonsei University, Seoul, Korea, in 1985 and 1987, respectively, and the Ph.D. degree in Computer Engineering from the University of Missouri, Columbia, in 1996.
From 1988 to 1991, he was with the Quality Assurance Center, Korea Telecom. From 1997 to 1999, he was a senior member of technical staff in Multimedia Technoloy Research Laboratory, Korea Telecom, where he developed a content based image search engine for the Web. He is currently the director of Image and Video Processing Research Division, Korea Telecom. His research interests encompasses many aspects of computer vision, pattern recognition. His current interests include the management and retrieval of audio-visual content, content personalization, content delivery network, and the standardization such as MPEG-7 and MPEG-21.