

ART의 위상 정보를 이용한 형태기반 영상 검색 방법

이종민^{a)}, 김회율^{a)‡}

A Shape Based Image Retrieval Method using Phase of ART

Jong-Min Lee^{a)} and Whoi-Yul Kim^{a)‡}

요 약

영상에 포함된 객체의 형태는 내용 기반 영상검색에 있어서 중요한 정보를 가지고 있기 때문에, 이를 이용하여 영상을 검색하는 방법들이 활발히 연구되어 왔다. 그중에서도 최근에 제안된 저니키 모멘트의 위상과 크기를 이용하는 회전불변 서술자(IZMD: Invariant Zernike moment descriptor)을 이용한 영상 검색 방법은 기존의 크기 정보만을 이용한 저니키 모멘트 서술자보다 높은 영상 검색 성능을 보인다. 본 논문에서는 IZMD를 이용한 방법 보다 향상된 영상 검색 성능을 얻기 위해서 ART(Angular Radial Transform)의 크기와 위상을 이용한 회전 불변 특징 서술자(IARTD: Invariant Angular Radial Transform Descriptor)와 이를 이용해서 영상을 정합하는 방법을 제안한다. IARTD는 ART 기저함수의 특성을 이용해서 정렬된 ART 계수의 위상과 크기로 구성된 특징벡터이다. 영상의 검색은 두 IARTD의 크기차이와 위상차이의 곱을 이용하여 정의된 거리 계산 방법을 이용해서 수행한다. MPEG-7 데이터셋을 이용한 실험 결과, 제안하는 방법의 평균 BEP(Bull's Eye performance)는 0.5806으로서, ART나 IZMD를 이용한 영상 검색 결과의 평균 BEP 0.3574, 0.4234보다 우수한 검색 성능을 제공하는 것을 확인하였다.

Abstract

Since shape of an object in an image carries important information in contents based image retrieval (CBIR), many shape description methods have been proposed to retrieve images using shape information. Among the existing shape based image retrieval methods, the method which employs invariant Zernike moment descriptor (IZMD) showed better performance compared to other methods which employ traditional Zernike moments descriptor in CBIR. In this paper, we propose a new image retrieval method which applies invariant angular radial transform descriptor (IARTD) to obtain higher performance than the method which employs IZMD in CBIR. IARTD is a rotationally invariant feature which consists of magnitudes and aligned phases of angular radial transform coefficients. To produce rotationally invariant phase coefficients, a phase correction scheme is performed while extracting the IARTD. The distance between two IARTDs is defined by combining the differences of the magnitudes and the aligned phases. Through the experiment using MPEG-7 shape dataset, the average bull's eye performance (BEP) of the proposed method is 0.5806 while the average BEPs of the existing methods which employ IZMD and traditional ART are 0.4234 and 0.3574, respectively.

Keyword : ART, phase, rotation invariant, image retrieval

a) 한양대학교 전자컴퓨터통신공학부

Department of Electronics and Computer Engineering

‡ 교신저자 : 김회율 (wykim@hanynag.ac.kr)

※ 본 연구는 한국전자통신연구원의 “깊이 영상 기반 실시간 3차원 동작 인식 알고리즘 개발” 연구의 일환으로 수행된 연구 결과임

· 접수일(2011년8월11일), 수정일(1차:2011년10월25일, 2차:12월7일), 게재확정일(2011년12월7일)

1. 서론

컴퓨터와 정보통신 기술의 발전 및 인터넷의 대중화에 의해 디지털 콘텐츠의 사용 및 그 정보량 또한 기하급수적으로 증가하고 있다. 디지털 콘텐츠 중에서도 사진이나 그

림과 같은 디지털 영상의 취급은 디지털 카메라와 컴퓨터를 이용한 영상 제작 프로그램들이 대중화되면서 더 이상 전문가들만의 영역이 아닐 정도로 보편화 되었다. 방대한 양의 영상정보속에서 사용자가 원하는 영상정보를 획득하는 것은 많은 시간과 노력을 필요로 하기 때문에 영상정보를 효율적으로 검색하는 방법에 대한 연구가 꾸준히 이루어져 왔다^[1]. 영상을 검색 하는 방법에는 크게 색인어를 이용한 텍스트 기반 영상 검색 방법과 영상의 특징을 이용한 내용기반 영상 검색 방법이 있으며, 내용기반 영상 검색 방법에 사용되는 영상의 특징으로는 주로 색상, 질감, 형태 등이 이용된다^[2].

영상의 내용적 특징중에서 형태 정보는 상표나 캐릭터와 관련된 영상을 검색하거나 정합하는데 있어서 높은 정확도를 제공하기 때문에, 객체의 형태를 효율적으로 서술하는 형태 서술자들이 그동안 다양한 연구를 통해서 제안되었다^[3]. 그 중에서도 저니키 모멘트(Zernike moments)는 기저 함수들간의 정보가 중복성이 없고, 기하(geometrical) 모멘트나 리젠드리(Legendre) 모멘트 같은 다른 모멘트들에 비해 잡음에 강인하고, 패턴 표현 능력이 뛰어나다^{[4][5]}. 저니키 모멘트를 이용해서 영상을 검색하거나 정합하는 많은 연구들은 주로 저니키 모멘트의 크기(magnitude)만을 영상의 특징으로 사용했었다^[6-8]. 저니키 모멘트의 크기를 이용하는 것은 회전에 불변하게 영상을 표현 할 수 있는 이점이 있지만, 반대로 회전정보 자체를 잃어버리는 단점도 있다. 따라서, 최근에 저니키 모멘트의 위상(phase)과 크기를 함께 이용해서 영상을 서술하고 정합하는 방법들이 제안되고 있다^[9-11]. 신호의 크기와 위상은 모두 영상을 서술하는데 중요한 정보를 담고 있기 때문에, 저니키 모멘트의 크기만을 사용하는 것 보다는 위상을 함께 사용하면 보다 정확하게 영상을 정합할 수 있다^{[9][12]}. 하지만, 회전된 영상들을 저니키 모멘트의 위상을 이용해서 정합하기 위해서는, 상대적인 회전각도를 추정하여 위상을 보정하거나 위상을 정렬하는 과정이 필요하다.

저니키 모멘트를 이용하여 두 영상간의 회전 각도를 추정하여 보정된 위상으로 영상을 정합하는 방법은 영상의 잡음에 강인하게 영상을 정합할 수 있기 때문에 높은 분별력을 요구하는 지역 특징점 정합(local feature matching)에

도 적용 가능하지만^{[10][11]}, 비교하는 영상들마다 회전 각도를 추정해야하기 때문에 영상 정합에 많은 계산량을 요구하는 단점이 있다. 영상간의 회전 각도를 추정하지 않고 저니키 모멘트의 위상을 정렬해서 회전에 불변한 특징값을 생성하여 영상을 정합하는 방법으로는, Shan 외 2명이 제안한 회전에 불변한 저니키 모멘트 서술자(IZMD: Invariant Zernike Moments Descriptor)를 이용한 방법이 있다^[9]. IZMD는 한 개의 영상에서 계산된 저니키 모멘트들 중의 하나의 위상을 기준으로 나머지 모멘트들의 위상을 정렬하여 특징벡터를 생성한다. IZMD를 이용한 영상정합은 회전각도를 추정하는 방법에 비해서 상대적으로 영상의 잡음에 취약하지만, 모멘트의 크기만을 사용하는 것보다 향상된 영상 정합성능을 얻을 수 있으며, 상대적으로 계산량이 적어서 빠른 영상정합이 가능하다.

위상정보를 이용하여 영상의 유사도를 비교하는 방법들이 최근 제안되고 있는 저니키 모멘트에 비해서, ART의 경우에는 위상정보를 이용하여 영상의 유사도를 비교하는 방법에 대한 연구는 아직 활발히 이루어지고 있지 않다. 하지만, ART가 모멘트의 크기 정보만을 이용했을 때 저니키 모멘트에 비해서 더 좋은 영상 검색 성능을 보이기 때문에^[13], 위상정보를 이용한 경우에도 더 좋은 영상 검색 성능을 제공할 것을 기대할 수 있다. 따라서, 본 논문에서는 [9]에서 제안된 IZMD를 이용한 영상 검색 방법보다 향상된 영상검색 성능을 얻기 위해서, ART(Angular Radial Transform)의 크기와 정렬된 위상들로 회전에 불변한 특징벡터(IARTD: Invariant ART Descriptor)를 생성하고 이를 이용해서 영상을 검색하는 방법을 제안한다. 제안하는 영상 검색 방법은 ART를 이용해서 IARTD를 생성하는 방법과 향상된 정합 결과를 제공하는 IARTD간의 거리 계산 방법을 포함한다.

본 논문의 이 후의 구성은 다음과 같다. II장에서 저니키 모멘트와 ART에 대해서 간략히 설명하고, III장에서는 저니키 모멘트의 위상정보를 이용한 회전불변 서술자 IZMD와 이를 이용한 영상 정합방법에 대해서 서술한다. IV장에서는 제안하는 IARTD를 이용한 영상 검색 방법을 자세히 설명하고, MPEG-7 데이터셋을 이용한 실험결과를 V장에 제시한 후, VI장에서 결론을 맺는다.

II. 저니키 모멘트와 ART

1. 저니키 모멘트

저니키 모멘트는 입력영상을 저니키 기저 함수에 투영하여 그 합을 계산함으로써 구해진다. 저니키 기저 함수는 반지름이 1인 원형 안에서 정의되기 때문에 입력 영상은 $x^2+y^2 \leq 1$ 인 기저함수에 투영되기 위해서 단위원 크기로 정규화된다. 차수가 n 이고 반복수가 m 인 저니키 모멘트 Z_{nm} 은 수식 (1)과 같이 정의 된다.

$$Z_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} f(x,y) V_{nm}^*(x,y) dx dy, \quad (1)$$

$f(x,y)$ 는 입력 영상을 의미하고, $V_{nm}(x,y)$ 는 저니키 기저 함수를 나타낸다. 기호 *는 복소 공액을 나타낸다. 차수 n 은 0을 포함한 양의 정수이며, 반복수 m 은 $n-|m|$ (짝수)와 $|m| \leq n$ 을 만족하는 정수이다. 저니키 기저함수 $V_{nm}(x,y)$ 는 실수 방사 다항식 $R_{nm}(\rho)$ 를 원형 주기함수와 결합하여 식 (2)와 같이 정의된다.

$$V_{nm}(x,y) = V_{nm}(\rho,\theta) = R_{nm}(\rho) \exp(jm\theta) \quad (2)$$

실수 방사 다항식 R_{nm} 은 아래와 같이 정의된다.

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s (n-s)! \rho^{n-2s}}{s! \binom{n+|m|}{1-s} \binom{n-|m|}{1-s}}. \quad (3)$$

저니키 모멘트를 이산 영상에 대해서 계산할 때는 아래와 같은 근사를 사용한다.

$$Z_{nm} = \frac{n+1}{\pi} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) V_{nm}^*(x,y). \quad (4)$$

저니키 모멘트를 이용한 기존의 영상 검색 방법들은 영상에서 계산된 저니키 모멘트들의 크기들만으로 구성된 특징벡터들의 유클리디언 거리를 영상의 유사도를 정의했다 [6-8].

2. ART

ART는 저니키 모멘트나 유사 저니키 모멘트보다 향상된 영상 정합 결과를 얻기 위해서 [11]에서 제안된 모멘트 기반 형상 서술자이다. ART의 기저 함수는 실수 방사 다항식이 코사인이나 사인함수로 정의된다. 또한, 차수가 홀수 일 때는 반복수도 홀수로 증가하고, 차수가 짝수 일 때는 반복수도 짝수로만 증가하는 저니키 모멘트와 달리 ART의 차수와 반복수는 서로 독립적이다. 차수가 n 이고 반복수가 p 인 ART 계수 F_{np} 는 수식 (5)와 같이 정의 된다.

$$F_{np} = \iint_{x^2+y^2 \leq 1} f(x,y) V_{np}^{ART*}(x,y) dx dy, \quad (5)$$

여기서 $V_{np}^{ART}(x,y)$ 는 ART 기저 함수를 나타내며, 식 (6)과 같이 정의된다.

$$V_{np}^{ART}(\rho,\theta) = A_p(\theta) R_n^{ART}(\rho). \quad (6)$$

$A_p(\theta)$ 는 ART의 원형 주기 함수로서 식 (7)과 같이 정의되고, 실수 방사 다항식 R_n^{ART} 은 식 (8)처럼 코사인이나 사인함수를 이용해서 두 가지 형태로 정의 될 수 있다.

$$A_p(\theta) = \frac{1}{2\pi} \exp(jp\theta), \quad (7)$$

$$\begin{aligned} ART-C: R_n^C(\rho) &= \begin{cases} 1 & n=0 \\ 2\cos(\pi n\rho) & n \neq 0 \end{cases} \\ ART-S: R_n^S(\rho) &= \begin{cases} 1 & n=0 \\ 2\sin(\pi n\rho) & n \neq 0 \end{cases} \end{aligned} \quad (8)$$

이 때, n 과 p 는 0을 포함한 양의 정수이다. 본 논문에서는 코사인 함수를 이용해서 생성된 ART 기저 함수를 이용하였다.

ART를 이용한 영상간의 유사도는 계산된 ART 계수들의 크기를 이용해서 구성된 특징벡터간의 유클리디언 거리를 이용해서 계산된다 [12]. MPEG-7에서는 ART를 이진 영상에 대해서만 정의하고 있지만, Ricard가 흑백영상이나 컬러영상 및 3D 객체에 대해서 ART 계수를 추출하는 방법

을 [14]에서 제안하였다.

III. IZMD를 이용한 영상 검색 방법

저니키 모멘트를 이용한 기존의 영상 검색 방법들보다 높은 영상 검색 성능을 얻기 위해서, Shan외 2명은 [9]에서 저니키 모멘트의 위상과 크기를 모두 포함하는 IZMD를 이용해서 영상 검색을 수행하는 방법을 제안하였다. IZMD를 이용한 영상 검색 방법은 저니키 모멘트를 추출하기전에 먼저 [6]에서 제안된 기하 모멘트를 이용한 영상의 정규화 과정을 통해서 저니키 모멘트 기저함수의 단위원에 특징벡터를 추출할 영상을 사상한다. 정규화 과정을 거친 영상의 저니키 모멘트는 이동과 크기변화에 불변하게 된다. 정규화 과정을 거친 영상에서 계산된 Z_{00} 과 Z_{11} 은 상수가 되기 때문에 IZMD 특징벡터를 구성하는데 사용되지 않는다. 계산된 저니키 모멘트들의 위상은 반복수 1에 해당하는 저니키 모멘트들의 위상을 이용해서 식(9)과 같이 정렬할 수 있다.

$$Z_{nm}(\rho, \theta) = Z_{nm}(\rho, \theta) e^{-jm\phi_{n0,1}}; \quad (9)$$

$$n_0 \in \left\{ 1, 3, 5, 7, \dots, \left(2 \cdot \left\lceil \frac{n_{\max}}{2} \right\rceil - 1 \right) \right\}.$$

여기서 Z 는 반복수 1에 해당하는 성분의 위상을 이용해서 전체 모멘트들의 위상이 정렬된 회전 불변 특징 값이다. 식 (9)에서 위상만을 분리하면 (10)처럼 나타낼 수 있다.

$$\phi'_{nm}(\rho, \theta) = \phi_{nm}(\rho, \theta) - m\phi_{n0,1}(\rho, \theta), \quad (10)$$

ϕ_{nm} 은 저니키 모멘트의 위상이고, $\phi'_{nm}(\rho, \theta)$ 는 정렬된 위상이다. 식 (10)을 이용하면 회전된 영상의 저니키 모멘트들의 위상 $\phi'^{rotated}_{nm}$ 을 정렬하는 식 (11)을 얻을 수 있다.

$$\phi'^{rotated}_{nm}(\rho, \theta) = \phi'^{rotated}_{nm}(\rho, \theta) - m\phi'^{rotated}_{n0,1}(\rho, \theta) = \phi'_{nm}(\rho, \theta) \quad (11)$$

위상의 정렬 기준이 되는 $\phi_{n0,1}$ 은 식 (9)에 정의된 것처럼, 반복수 1에 해당하는 기저함수에 의해 계산된 모멘트들의

위상 중에서 어떤 것이라도 사용 가능하지만, 정규화 과정을 거친 영상의 $Z_{1,1}$ 은 0 이되기 때문에 사용할 수 없다. [7]에서는 잡음에 강인하게 위상을 정렬하기 위해서, 사용 가능한 저니키 모멘트중에서 가장 낮은 차수인 $Z_{3,1}$ 의 위상 성분을 기준으로 나머지 모멘트들의 위상을 정렬했다. IZMD 특징벡터는 계산된 모멘트들의 크기와 식 (11)을 이용해서 정렬된 모멘트들의 위상으로 구성된다. IZMD는 반복수 0에 해당하는 모멘트들의 위상과 크기를 특징벡터에 포함하고 있는데, 반복수가 0인 모멘트들은 기저함수의 허수성분이 모두 0이기 때문에 영상의 특징 표현에 의미있는 위상 정보를 가지고 있지 않다. 따라서 반복수 0의 모멘트들을 특징값으로 사용하는 것은 바람직 하지 않다.

IZMD를 이용한 두 영상 간의 정합은 영상으로부터 계산된 두 IZMD의 각 성분에서 크기 성분들의 차이의 합과 위상 성분들의 차이의 합을 따로 계산한 뒤 가중합을 계산하는 식 (12)로 정의된다.

$$D = w_{ang} D_{ang} + w_{mag} D_{mag},$$

$$D_{mag} = \frac{1}{N} \sqrt{\sum_{i=1}^N \left(\frac{(\phi_i^{query} - \phi_i^{stored})}{\pi} \right)^2}, \quad (12)$$

$$D_{ang} = \frac{1}{N} \sqrt{\sum_{i=1}^N \left(\frac{|Z_i^{query}| - |Z_i^{stored}|}{\max(|Z_i^{query}|, |Z_i^{stored}|)} \right)^2}.$$

이 때, N 은 IZMD를 구성하는데 사용된 저니키 모멘트의 개수이며, i 는 IZMD를 생성하는데 사용된 저니키 모멘트의 인덱스이다. *query*와 *stored*는 각각 질의 영상과 DB에 저장된 영상을 의미한다. w_{ang} 와 w_{mag} 는 가중치로서 $w_{ang} + w_{mag} = 1$ 을 만족한다.

IV. 제안하는 영상 검색 방법: IARTD

본 논문에서는 IZMD나 기존의 ARTD 보다 높은 영상 검색 성능을 얻기 위해서, ART 계수의 위상정보를 이용하여 영상을 서술하고, 영상간의 유사도를 비교하는 IARTD를 제안한다. IARTD를 생성하는데 이용되는 ART는 저니

키 모멘트에 비해서 영상의 회전방향 변화를 중점적으로 서술하는 기저함수 집합을 생성할 수 있으며, 이는 사람이 형태정보를 인지하는 함에 있어서 방사방향의 정보보다 회전방향의 정보에 더 의존하는 것과 유사하다. 이러한 특성을 갖는 기저함수 집합의 ART 계수는 유사한 개수의 저니키 모멘트에 비해서 영상간 정합의 정확도가 높다^[15]. ART는 저니키모멘트와 동일하게 단위원에서 정의되는 기저함수를 이용하여 계산되기 때문에, 먼저 입력영상은 [3]에서 제안한 방법을 이용해서 정규화된다. 정규화된 원영상과 회전된 영상의 ART의 계수들의 위상 성분간의 차이 θ_{np} 는 식 (13)과 같이 원영상의 ART 계수를 회전된 영상의 ART 계수로 나눈 뒤 그 편각을 계산하여 구할 수 있다.

$$\theta_{np} \equiv \arg\left(\frac{F_{np}^{rotated}}{F_{np}}\right) = p\alpha, 0 \leq \theta_{np} \leq 2p\pi, \quad (13)$$

여기서 $F_{np}^{rotated}$ 은 회전된 영상의 ART 계수이고, α 는 회전각도이다. 식 (13)은 ART 계수의 위상과 회전각도 α 를 이용해서 식 (14)처럼 표현할 수도 있다. 이 때, Φ_{np} 은 원본 영상과 회전된 영상의 ART 계수간의 위상 차이이고, ϕ_{np} 은 ART 계수의 위상이다.

$$\begin{aligned} \Phi_{np} &= (\phi_{np}^{rotated} - \phi_{np}) \bmod(2\pi) \\ &= (p\alpha) \bmod(2\pi), 0 \leq \Phi_{np} \leq 2\pi, \end{aligned} \quad (14)$$

ART는 기저함수와 저니키 모멘트의 기저함수는 유사하게 정의되기 때문에, 계수들의 위상도 유사하다. 따라서, IZMD의 위상정렬에 관한 식 (10)과 식 (11)을 확장하여 ART 계수들의 위상이 회전에 불변한 특징값을 갖도록 정렬할 수 있다. 다만, 저니키 모멘트는 홀수의 차수에서만 반복수 1에 해당하는 계수가 정의되는 반면, ART는 모든 차수에서 반복수 1에 해당하는 계수를 얻을 수 있다. 그러므로, IZMD의 위상 정보를 정렬하는 식 (9)를 IARTD의 경우에는 식 (15)처럼 정의한다.

$$F'_{np}(\rho, \theta) = F_{np}(\rho, \theta) e^{-jp\phi_{n,1}}. \quad (15)$$

여기서 F' 은 위상이 정렬된 ART 계수이다.

IARTD는 식 (15)에 의해서 각 차수마다 반복수 1에 해당하는 계수를 기준으로 정렬된 위상과 크기로 구성된 특징벡터이다. IARTD는 유효한 위상정보를 가지고 있지 않은 반복수 0에 해당하는 계수들은 특징벡터에 포함하지 않는다. 각 차수의 반복수 1에 해당하는 계수들도 위상정보 정렬을 수행한 후에는 위상정보가 0이 되기 때문에 특징벡터에 포함하지 않는다.

ART 계수들의 크기와 정렬된 위상으로 구성된 IARTD의 특징벡터를 이용해서 효과적으로 영상의 정합을 수행하기 위해서 본 논문에서는 두 개의 IARTD 간의 거리를 식 (16)으로 정의한다.

$$\begin{aligned} D_{IARTD} &= \sum_n \sum_{p \geq 2} \| |F'_{np}{}^{query}| - |F'_{np}{}^{stored}| \| \times \phi_{np}^{diff}, \\ \phi_{np}^{diff} &= \frac{|\phi_{np}^{query} - \phi_{np}^{stored}|}{\pi}; 0 \leq \phi_{np}^{diff} \leq 1. \end{aligned} \quad (16)$$

여기서 $|\phi_{nm}^{query} - \phi_{nm}^{stored}|$ 는 위상 공간에서의 차이의 절대값이기 때문에 $[0, \pi]$ 사이의 값을 갖는다.

V. 실험 결과

실험에서는 IARTD의 최적 파라미터를 결정하기 위한 실험과, 제안하는 IARTD를 이용한 영상 검색 방법의 성능을 검증하기 위해 IZMD를 이용한 영상 검색 방법과의 성능 비교를 수행하였다. 각 서술자를 이용한 영상 검색 결과는 질의 영상과 데이터셋 내의 영상들 사이의 유사도를 각 서술자의 거리함수 식 (12)와 식 (16)을 이용해서 계산하고, 계산된 거리에 따라 오름차순으로 정렬된 영상들로 제시된다. IARTD와 IZMD를 계산하기 위한 입력영상은 [6]에서 제안한 방법을 이용해서 101×101 로 정규화 하였다. IZMD의 경우, [9]에서 제안된대로 차수 10까지의 저니키 모멘트를 이용하여 68차원의 특징벡터를 생성하였으며, 각 서술자의 기저함수들의 단위원은 데이터셋의 영상에 내접하는 원으로 사상되었다.

1. 실험 영상

실험은 MPEG-7의 region shape DB CE-1의 A1, A2, B 데이터셋과 B 데이터셋에 세 단계의 임펄스 잡음을 가하여 생성한 영상들을 이용하여 수행하였다. CE-1 데이터셋은 70개의 대표 영상들에 대해서 크기 변화(A1), 회전(A2), 변형(B)이 가해진 세 가지 데이터셋으로 구성되어 있다. 그림 1은 대표영상들의 일부이다. 각 데이터셋의 자세한 구성은 다음과 같다.

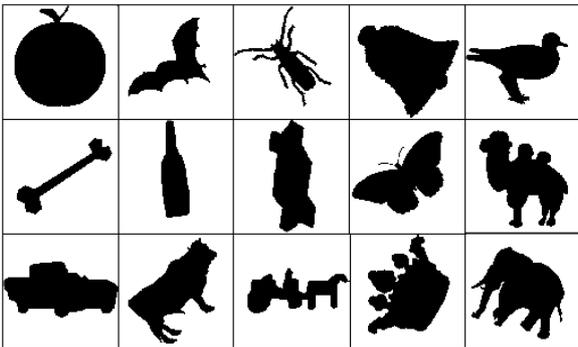


그림 1. MPEG7 CE-1 A1, A2, B 데이터셋의 각 그룹의 대표영상
Fig. 1. Sample images of MPEG7 CE-1 A1, A2 and B dataset

1.1 크기 변화에 대한 데이터셋

CE-1 A1 데이터셋은 70개의 그룹, 총 420장의 영상이 포함되어 있다. 각 그룹은 1개의 대표영상에 대해서 5 단계로 (0.1, 0.2, 0.25, 0.3, 2) 크기가 변화된 영상으로 구성되어 있다.

1.2 회전에 대한 영상

CE-1 A2 데이터셋은 70개의 그룹, 총 420장의 영상이 포함되어 있다. 각 그룹은 1개의 대표영상에 대해서 5 단계로 (9°, 36°, 45°, 90°, 150°) 회전된 영상으로 구성되어 있다.

1.3 변형에 대한 데이터셋

CE-1 B 데이터셋은 70개의 그룹, 총 1400장의 영상이 포함되어 있다. 각 그룹은 1개의 대표영상과 부분적인 변형을 가한 19개의 변형된 영상으로 구성되어 있다. 그림 2는

변형된 영상들의 예이다. 각각의 2 줄은 같은 그룹에 속한 영상들이다.

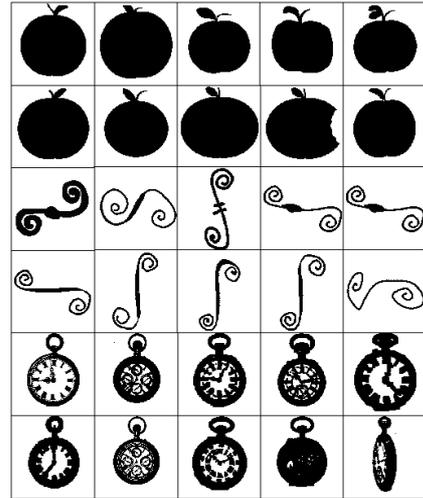


그림 2. 변형에 대한 데이터셋의 예 (CE-1 B)
Fig. 2. Sample images of deformed image dataset (CE-1 B)

1.4 잡음에 대한 데이터셋

영상 검색 방법들의 잡음에 대한 강인성을 비교하기 위해서 CE-1 B 데이터 셋에 3단계의 (해상도 대비 3%, 5%, 7%) 임펄스 잡음을 가하여 3개의 데이터 셋을 생성하였다.

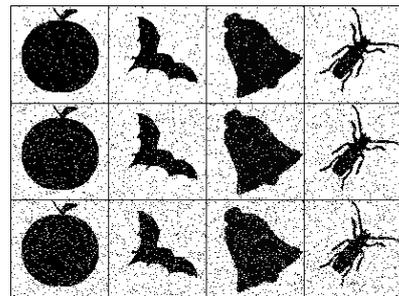


그림 3. 잡음에 대한 데이터셋 (위: 3%, 중간: 5%, 아래:7%)
Fig. 3. Sample images with impulse noise under different levels (top: 3%, middle:5%, bottom: 7%)

2. 성능 비교 기준

본 논문에서 제안한 IARTD와 IZMD의 영상 검색 성능을 비교하기 위한 기준으로 BEP (Bull's Eye Performance)

를 사용하였다. BEP는 질의 영상에 대한 검색 결과에서, 질의 영상과 같은 그룹에 속하는 영상의 갯수와 그 갯수의 두 배 범위 내에 포함된 질의영상과 동일한 그룹에 속하는 영상들의 갯수의 비율이다. 따라서, BEP는 0에서 1사이 값을 가지며 높을수록 우수한 검색 성능을 나타낸다. 각 검색 방법의 성능 비교는 데이터셋 내의 모든 영상들을 이용해서 검색을 수행하여 얻어진 BEP들의 평균값을 사용하였다.

3. IARTD의 파라미터 결정

ART를 이용해서 영상의 특징을 효율적으로 서술기 위해서는 영상의 해상도에 적합한 차수 n 과 반복수 p 의 최대값을 결정해야 한다. 최대 차수와 최대 반복수는 ART를 이용한 영상 서술자의 분별력 및 계산의 효율성과 관련된다. 다른 검색 방법과 IARTD의 성능을 비교하기 전에, IARTD의 최대 차수와 최대 반복수를 결정하기 위해서 두 파라미터를 변화시켜가면서 검색결과를 비교하였다. 먼저 최대 차수를 결정하기 위해 최대 반복수를 10으로 고정시키고 최대 차수를 1에서 4까지 바꾸어가며 CE-1 B셋을 대해서 검색을 수행한 결과를 비교하였다. 그림 4는 최대차수에 따른 검색 결과들의 평균 BEP의 그래프이다.

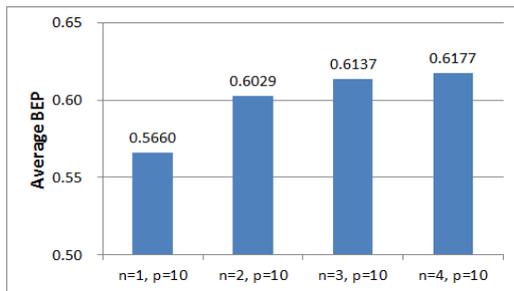


그림 4. IARTD의 최대 차수에 따른 평균 BEP
Fig. 4. Average BEPs of IARTD of different max orders

그래프에 나타난 것처럼, 전체 검색 성능의 변화는 크지 않지만, 최대 차수가 커질수록 검색성능이 향상 되는 것을 볼 수 있다. 최대 차수가 3과 4인 경우는 최대 차수가 1에서 2로 바뀔 때 보다 검색 성능의 향상되는 정도가 상대적으로 작다. 차수가 1 늘어날 때 마다 IARTD를 생성하는데 필요

한 ART 계수는 9개씩 증가하며 이는 IARTD를 계산하고, 영상을 정합하는데 더 많은 시간을 요구하게 된다. 표 1은 최대 반복수를 10으로 고정했을 때, 차수의 변화시키면서 한 장의 영상에서 IARTD를 추출하는 평균 계산시간을 측정한 것이다. ART의 계산은 Intel(R) Core(TM)2 Quad 2.4GHz CPU를 이용한 컴퓨터를 이용하여 수행하였으며, 10000장의 영상을 이용해서 평균을 계산하였다.

표 1. 최대 차수에 따른 IARTD의 계산 시간
Table 1. Computational times of IARTDs for different n_{max}

최대 차수 (n_{max})	1	2	3	4
실행시간 (μs)	6245.25	9653.30	13702.22	17539.58

차수가 1 증가할 때마다, 평균 3700 μs 정도의 계산시간이 증가하는 것을 알 수 있다. 하지만, 검색성능은 최대차수가 1에서 2로 증가할 때는 3.68%증가하지만, 최대차수가 2에서 3으로 증가할 때는 1.08%가 증가하는데 그친다. 즉 계산시간의 증가량은 일정하지만, 검색성능의 증가량은 약 3분의 1수준으로 감소한다. 따라서, 차원이 커지면서 계산 시간이 늘어나는 것과 검색 성능의 향상 정도를 고려했을 때, 최대 차수를 2로 하는 것이 효율적임을 알 수 있다.

계속해서 IARTD의 최대 반복수를 결정하기 위해서, IARTD의 최대 차수를 2로 고정한 후, 최대 반복수를 8에서 12까지 변화 시켜가면서 CE-1 B 셋에 대한 검색결과들의 평균 BEP를 비교하였다. 그림 5는 각 경우에 대해서 계산된 평균 BEP의 그래프이다.

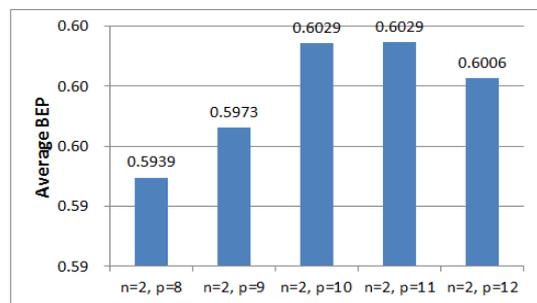


그림 5. IARTD의 최대 반복수에 따른 평균 BEP
Fig. 5. Average BEPs of IARTD of different max repetitions

최대 반복수가 10이 될 때까지는 BEP가 꾸준히 증가하지만 11에서 BEP의 개선이 눈에 띄게 둔화되더니 최대 반복수가 12일 때는 오히려 감소하였다. 이는 지름 101픽셀 크기의 단위원으로 ART 기저함수를 양자화 함에 있어서, 반복수가 12이상이면 기저함수의 양자화 오류에 의해서 ART 계수가 부정확해지기 때문으로 사료된다. MPEG-7에서도 101×101 영상의 경우 최대 반복수를 10으로 할 것을 권고하고 있다^[15].

위의 실험결과에 기반해서 이후의 실험에서는 IARTD의 최대 차수는 2, 최대 반복수를 10으로 고정하여 다른 검색 방법과의 검색 성능 비교를 수행한다. 따라서, IARTD를 구성하기 위해 사용되는 ART 계수의 수는 27개이며 특징벡터는 54차원으로서 [9]에서 사용된 IZMD의 68 차원보다 작다. 표 1과 같은 환경에서 1개의 영상에서 IZMD를 계산하는 시간의 평균을 구한 결과 11759.72 μ s 가 소요됨을 확인하였다. 표 1의 최대 차수 2, 최대 반복수 10인 IARTD의 평균 계산시간은 9653.30 μ s 이므로, 제안하는 IARTD가 IZMD에 비해서 계산시간이 짧은 것을 알 수 있다.

4. 위상 정렬 방법에 따른 검색 성능 비교

IARTD를 이용한 영상 검색 방법의 성능을 다른 방법들과 비교하기 전에, 본 논문에서 제안한 IARTD의 위상 정렬 방법의 효율성을 검증하는 실험을 수행하였다. 본 논문에서 사용된 IARTD의 최대 차수가 2 이므로, 위상 정렬의 기준으로 사용할 수 있는 반복수 1의 계수는 3개이다. 따라서, CE-1 B 데이터셋과 여기에 임펄스 잡음을 첨가한 3개의 데이터셋에 대해서, 세 개중 하나의 모멘트의 위상을 기준으로 위상을 정렬하여 생성한 IARTD와 본논문에서 제안한 방법으로 위상을 정렬한 IARTD에 의한 영상 검색 결과를 비교하였다. 그림 6은 위상 정렬 방법에 따른 영상 검색 결과의 그래프이다. 한 개의 위상 성분만을 이용하여 위상성분을 정렬하는 경우, 가장 낮은 차수인 0의 모멘트를 기준으로 하는 것이 가장 높은 검색 결과를 보였다. 특히 임펄스 잡음이 첨가된 영상에서는 검색 성능의 차이가 현저했으며, 이는 낮은 차수의 계수가 잡음에 더 강인하기 때문이다. 하지만, 제안하는 방법과 차수 0의 계수를 기준으

로 정렬 했을 때의 검색 성능을 비교했을 때는, 제안하는 방법을 사용했을 때 더 좋은 검색 결과를 얻을 수 있었다.

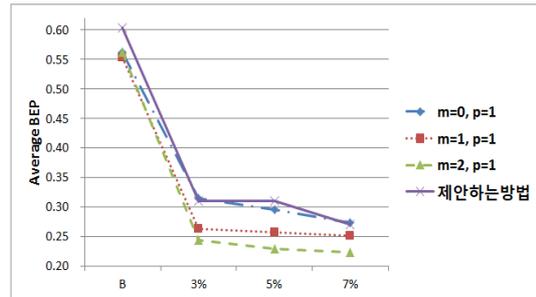


그림 6. 위상 정렬 기준별 IARTD 평균 BEP
Fig. 6. Average BEPs of IARTDs performed phase correction by different components

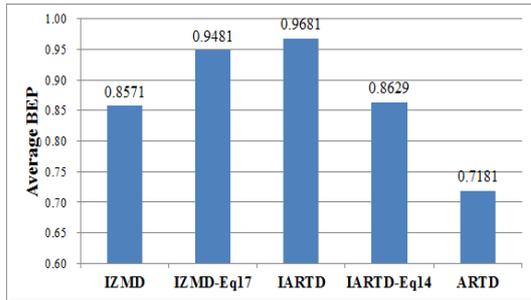
5. IARTD와 IZMD 검색 성능 비교

검색 성능 비교에서는 검색 결과의 변화가 IARTD를 이용한 영상 서술에 의한 것인지, IARTD의 거리 계산 방법에 의한 것인지를 분석하기 위해, 제안하는 방법에 의한 검색 결과를 IARTD에 IZMD의 거리 계산식 (14)를 적용한 검색 결과(IARTD-14) 및 ART의 크기만을 이용하는 기존의 ARTD의 검색 결과와 비교하였다. 또한, IZMD를 사용한 방법과의 검색 성능을 비교하기 위해서 IZMD에 의한 검색 결과와 본 논문에서 제안된 거리 계산식 (16)을 적용한 검색 결과(IZMD-16)를 함께 제시한다.

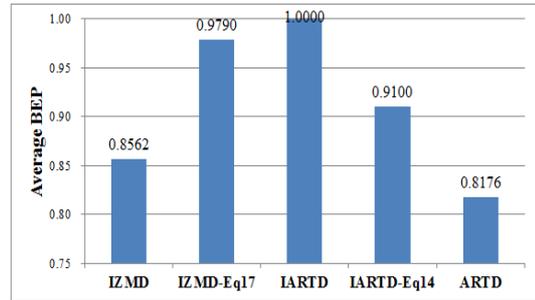
5.1 크기 변화, 회전, 변형에 대한 데이터셋

그림 7은 각 각 CE-A1, A2, B 데이터셋에 대해서 5개의 검색 방법을 이용해서 검색을 실시하여 얻은 검색 결과들의 평균 BEP를 그래프로 나타낸 것이다.

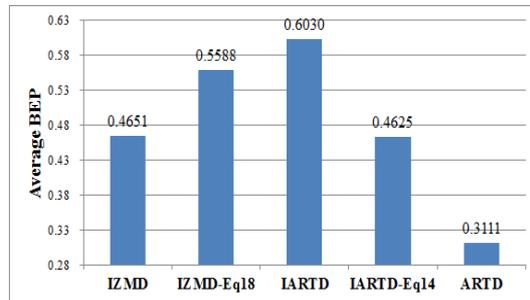
세 가지 데이터 셋에 대해서 검색을 수행한 결과, IARTD가 가장 우수한 검색 결과를 제공하는 것을 그래프에서 확인할 수 있다. 전체적으로 데이터셋에 따라, 다섯 가지 검색 방법에 성능의 변화는 있으나, 상대적인 검색 성능의 우위는 유사하게 유지 되었다. 각 각 동일한 특징벡터간 거리 계산 방법을 사용한 IMZD 와 IARTD-14, IARTD와 IZMD-16에 의한 검색결과를 비교해보면, 동일한 거리 계



(a) A1 데이터셋에 대한 영상 검색 성능



(b) A2 데이터셋에 대한 영상 검색 성능



(c) B 데이터셋에 대한 영상 검색 성능

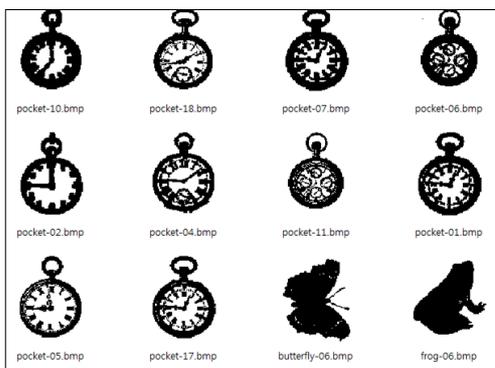
그림 7. 실험 데이터셋에 CE-1 A1, A2, B에 대한 영상 검색 방법들의 평균 BEP

Fig. 7. Average BEPs of each image retrieval method on experimental dataset CE-1 A1, A2 and B

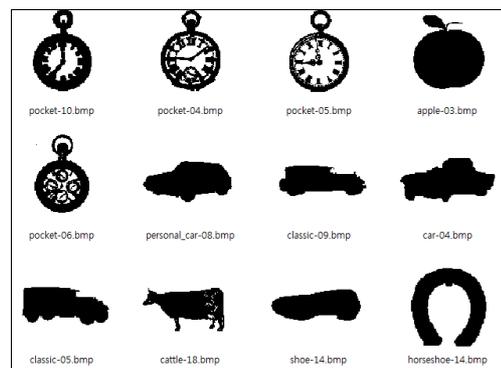
산 방법에 대해서 ARTD가 ZMD보다 우수한 검색 성능을 보이는 것을 알 수 있다. 또한, IZMD와 IZMD-16, IARTD와 IARTD-14를 비교 하면, 동일한 형태 서술자를 이용하는 경우 제안하는 특징벡터의 거리 계산 방법 식 (16)이 [6]

에서 제안된 특징값 거리 계산 식(14)보다 향상된 검색 결과를 제공하는 것을 알 수 있다. 전통적인 ARTD는 가장 낮은 검색 성능을 보였다.

그림 8은 변형된 데이터셋에 대해서 IARTD와 IZMD를



(a) IARTD를 이용한 검색 결과 (BEP: 0.8954)



(b) IZMD를 이용한 검색 결과 (BEP: 0.7371)

그림 8. 변형에 대한 데이터셋(CE-1 B)에 대한 검색 결과의 예 (상위 12개)

Fig. 8. Examples of image retrieval results of IARTD and IZMD on deformed image dataset (CE-1 B)

이용해서 영상검색을 실시한 결과의 일부를 예로 보여주고 있다. 각 결과에서 왼쪽 상단의 영상은 질의 영상이다.

5.2 잡음에 대한 데이터셋

그림 9는 CE-1 B셋에 3단계의 임펄스 잡음을 추가하여 생성한 잡음에 대한 3개의 데이터셋에 대한 다섯 가지 검색 방법에 의한 검색 결과를 그래프로 나타낸 것이다.

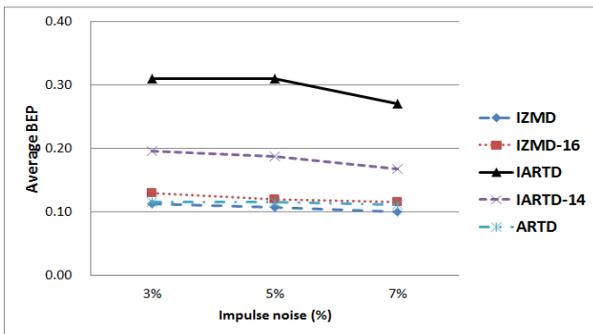


그림 9. 잡음 데이터셋에 대한 평균 BEP

Fig. 9. Average BEPs of each method on noisy dataset

5개의 검색 방법 모두 잡음이 추가 되지 않은 CE-1 B 데이터셋에 대한 검색 결과보다 상대적으로 저조한 검색 결과를 제공했다. 하지만, IARTD는 나머지 네 가지 검색 방법보다 평균 0.1 BEP 이상 높은 검색 성능을 보였다. IZMD와 본 논문에서 제안한 거리 계산 방법을 이용한 검색 방법 IZMD-16은 앞의 세 개의 데이터셋에서는 IZMD와 IARTD-14보다 더 높은 검색 성능을 보였으나, 잡음에 대한 데이터셋들에서는 거의 비슷한 검색 결과를 보였다. 따라서 임펄스 잡음과 같은 픽셀 수준의 랜덤 노이즈에 대해서는 ART가 저니키 모멘트보다 상대적으로 강인하게 형태 정보를 서술하는 것을 알 수 있다.

VI. 결론

본 논문에서는 형태 기반 영상 검색을 위해서, ART의 위상 성분을 이용하면서도 회전에 불변한 IARTD를 이용한 영상 검색 방법을 제안하였다. 제안하는 영상 검색 방법은

기존의 ARTD나 IZMD의 보다 향상된 영상 검색 성능을 얻기 위한 특징 벡터 생성 방법과 특징 벡터 간의 거리 계산 방법을 포함한다. 실험에서는 제안하는 방법의 성능을 검증하기 위해서, MPEG-7 데이터셋을 이용해서 기존의 형태 형태기반 영상 검색 방법들과 영상 검색 성능을 BEP를 기준으로 비교하였다. 실험결과, 제안하는 영상 검색 방법의 검색 성능은 평균 0.5806 BEP로서, 실험에 사용된 방법들 중에서 가장 높은 검색 성능을 보였으며, 이는 IZMD를 사용한 방법의 평균 0.4234 BEP 보다 약 0.16 BEP 높았다. 특히, IARTD는 특징벡터의 차원이 IZMD에 비해 낮으면서도, 상대적으로 좋은 검색 결과를 제공하였다. 또한, 동일한 특징벡터를 사용할 때 본 논문에서 제안하는 특징값 간의 거리 계산 방법이 [6]에서 제안된 거리 계산 방법보다 향상된 영상 정합 결과를 제공하는 것을 확인할 수 있었다.

참고 문헌

- [1] S. Chang, J. Smith, M. Beigi and A. Benitez, "Visual Information Retrieval from Large Distributed Online Repositories," *Communications of ACM*, Vol. 12, pp. 12-20, 1997.
- [2] Y. Rui, T. Huang, and S. Chang, Image retrieval: current techniques, promising directions and open issues, *J. of Visual Communication and Image Representation*, vol. 10, no. 4, 39-62, 1999.
- [3] Y. Mingqiang, K. Kidiyo, and R. Joseph, (2008) "A Survey of Shape Feature Extraction Techniques," on *Pattern Recognition Techniques, Technology and Applications*, Vienna: i-Tech, pp.626.
- [4] R. J. Prokop and A. P. Reeves, "A survey of moment-based techniques for unoccluded object representation and recognition," *Graphical Models and Image Processing*, vol. 54, no. 5, pp. 438-460, Sep. 1992.
- [5] C. H. The and R. T. Chin, "On image analysis by the method of moments," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 10, no. 4, pp. 496-513, July. 1998.
- [6] A. Khotnazard and Y. H. Hong, "Invariant image recognition by Zernike moments," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 489-497, May. 1990.
- [7] M. Teague, "Image analysis via the general theory of moments," *Journal of the Optical Society of America*, Vo. 70. pp. 920-930, Aug. 1980.
- [8] Y. S. Kim and W. Y. Kim, "Content-based trademark retrieval system using visually salient feature," *Journal of Image and Vision Computing*, vol. 16, pp. 931-939, Aug. 1998
- [9] Shan Li, Moon-Chuen Lee, and Chi-Man Pun, "Complex Zernike Moments Features for Shape-Based Image Retrieval," *IEEE Trans. on Systems, Man, And Cybernetics-Part A: Sstems and Humans*, vol. 3,

- no. 1, Jan. 2009
- [10] J. Revaud et al., "Improving Zernike moments comparison for optimal similarity and rotation angle retrieval," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 31, no. 4, pp. 627-637, Apr. 2009.
- [11] Z. Chen and S.K. Sun, "A Zernike moment phase-based descriptor for local image representation and matching," IEEE Trans. Image processing, vol. 19, no. 1, Jan 2010.
- [12] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," Proc. IEEE, vol. 69, no. 5, pp. 529 - 550, 1981.
- [13] S. Jeannin, "Mpeg-7 Visual part of eXperimentation Model Version 9.0," in ISO/IEC JTC1/SC29/WG11/N3914, 55th Mpeg Meeting, Pisa, Italia, Jan. 2001.
- [14] J Ricard, D Coeurjolly, A. Baskurt, "Generalization of angular radial transform for 2D and 3D shape retrieval," Pattern Recognition Letters, vol. 26, no. 14, pp. 2174-2186, Oct. 2005.
- [15] W. Y. Kim and Y. S. Kim, "A new region-based shape descirptor: The ART (Angular Radial Transform) Descriptor," ISO/IEC MPEG99/M5472, Maui, Dec. 1999.

저 자 소 개



이 종 민

- 1998년 2월 : 경원대학교 토목공학과 공학사
- 2003년 8월 : 한양대학교 정보통신대학원 석사
- 2004년 3월 ~ 현재 : 한양대학교 전자컴퓨터통신공학과 박사과정
- 주관심분야 : 형태서술자, 지역특지 정합, 객체 인식, 머신 비전



김 회 울

- 1980년 : 한양대학교 전자공학과 공학사
- 1983년 : Pennsylvania State University 전자공학과 공학석사
- 1989년 : Purdue University 전자공학과 공학박사
- 1989년 ~ 1994년 : University of Texas 조교수
- 1994년 ~ 현재 : 한양대학교 전자컴퓨터통신공학부 정교수
- 주관심분야 : 영상처리, 컴퓨터비전, 패턴인식, 머신비전, MPEG-7