

영상의 전체 및 지역 형태 정보를 이용한 내용 기반 영상 검색 기법

한두진, 박호연, 김현술, 박상희
연세대학교 전기·전자공학과

Content-Based Image Retrieval Using Global and Local shape information

Doo-Jin Han, Ho-Yeon Park, Hyun-Sool Kim, Sang-Hui Park
Dept. of Electrical and Electronic Engineering, Yonsei University

Abstract - 본 논문에서는 영상의 형태 정보를 이용하여 내용 기반 영상 검색을 수행할 수 있는 방법으로, 질의(query) 영상의 의사 저나이크 모멘트에서 영상내의 물체 형태에 대한 기여도가 가장 큰 모멘트를 추출하여 영상 전체의 형태 정보를 대표하는 특징벡터로 정하여 영상 검색을 수행하는 방법과, 영상의 인터레스트 포인트에서 미분 불변치 벡터와 위치 특성 벡터를 계산하여 영상의 지역 형태 정보를 대표하는 특징벡터로 정하여 영상 검색을 수행하는 방법, 그리고 두 가지 방법을 모두 고려하여 영상 검색을 수행하는 방법을 제시한다. 트레이드마크 영상 데이터베이스에 대해 영상 검색을 수행하여 기존의 영상 검색 방법과의 비교를 통하여 제안한 방법의 우수함을 보인다.

1. 서 론

내용 기반 영상 검색 기술은 영상 데이터베이스 구축 시 요구되는 연구분야로서 질의 자료로 영상 자체를 제시하면 질의 자료를 분석하여 색상, 형태, 텍스처(Texture) 등의 정보를 추출하여 영상 데이터베이스 상에서 가장 유사한 내용을 제시하는 기술이다. 요즘 많은 정보들이 멀티미디어화, 디지털화 되고 있으며 또한 저장 장치 기술의 발전으로 인하여 많은 자료들의 데이터베이스가 구축되고 있다. 이러한 광대한 영상 데이터베이스 상에서 원하는 영상을 일일이 찾아내는 일은 많은 시간과 노동을 필요로 한다. 따라서 원하는 영상과 유사한 영상이나 특징 정보를 질의 자료로 제시하고 자동적으로 질의 자료와 유사한 영상들을 순서대로 출력하면 검색 시간을 단축할 수 있으며 효율적인 검색이 가능하다.

본 연구에서는 인간의 시각적 인식 메커니즘과 유사하게 영상내에 포함된 객체의 전체 형태 및 지역의 형태 정보를 이용하였다. 전체 형태 정보를 이용하는 방식으로 영상의 대표 저나이크 모멘트를 특징벡터로 정하여 영상검색을 수행하는 방법을 사용하고, 지역 형태 정보를 이용하는 방식으로 미분 불변치 벡터와 위치 특성 벡터를 구성되는 특징벡터를 계산하여 영상검색을 수행하는 방법을 사용하고, 두 방식을 복합한 내용 기반 영상 검색 기법을 제안하고자 한다.

2. 전체 형태 정보를 이용한 내용 기반 영상 검색

데이터베이스에서 영상 검색을 수행하기 위해, 의사 저나이크 모멘트 중에서 영상내의 물체의 형태에 대한 기여도가 가장 큰 모멘트 즉, 시각적으로 두드러진 특징을 나타내는 대표 저나이크 모멘트만을 이용하는 방식으로 이를 향상된 저나이크 모멘트 방식(IZM)이라고 정의하였다.

2.1 의사 저나이크 모멘트 방식

의사 저나이크 모멘트(Pseudo Zernike Moments)는 단위원의 내부에서 완전한 직교집합을 형성하는 저나이크 다항식으로부터 유도된다[1]. 이들 다항식 집합은 $V_{nm}(x, y)$ 에 의해서 나타내어진다.

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

여기서, $n : 0$ 과 양의 정수

$m : |m| \leq n$ 조건을 만족하는 정수

$\rho, \theta :$ 중심까지의 거리와 시계반대방향의 각도

$R_{nm}(\rho) :$ 방사형 다항식

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

디지털 영상에 대한 n, m 차의 의사 저나이크 모멘트는 정규 모멘트와 비슷하게 다음과 같이 정의된다.

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) V_{nm}(x, y) x^2 + y^2 \leq 1$$

여기서, $f(x, y)$ 는 정규화된 이진 영상함수

의사 저나이크 모멘트는 회전에 불변하는 특성을 가지고 있고, 이동 및 크기변화의 경우에도 정규모멘트를 이용한 정규화과정으로 전처리 함으로써 불변하는 특성을 유지한다.

2.2 의사 저나이크 모멘트의 차수에 따른 시각적 효과
의사 저나이크 모멘트는 방사형 다항식과 위상형의 꼴로 계산된다. n 과 m 의 차가 짜우록 방사형 다항식은 고차항을 포함하게 되어 고주파성분을 나타낸다. 그래서 영상의 의사 저나이크 모멘트를 구했을 때, $n-m$ 의 차가 큰 모멘트가 큰 값을 가지게 되면 그 영상은 방사형으로 고주파 성분을 가지고 있고, 이와 반대로 적은 값의 $n-m$ 의 차가 큰 값의 모멘트를 갖게 되면 그 영상은 방사형으로 저주파 성분을 가지고 있다.

모멘트의 위상성분은 m 의 값에 따라 원형대칭을 나타낸다. 즉, 모멘트의 크기가 가장 큰 경우의 m 값에 따라, m 원형대칭의 영상에 해당된다. 예를 들어 m 이 2인 경우에는 가늘고 긴 형태의 영상이고, m 이 3인 경우에는 삼각형 형태, m 이 4인 경우에는 사각형 형태, m 이 5인 경우에는 오각형 형태의 영상이 된다[2].

2.3 대표 저나이크 모멘트 추출

항상된 저나이크 모멘트(IZM) 방식은 영상을 대표하는 특징 벡터로 의사 저나이크 모멘트 중에서 영상내의 물체의 형태에 대한 기여도가 가장 큰 경우의 모멘트를 사용한다. 다른 모멘트의 값은 영상의 형태를 구성하는데 기여도가 적으므로 무시한다. 질의 영상의 n, m 에 해당하는 의사 저나이크 모멘트인 Z_{nm} 변수의 질의 영상에 대한 기여도의 척도를 나타내기 위해 $D(n, m)$ 으로 정의 하면, $D(n, m) = P(Z_{nm} \geq z_{nm}) = 1 - F(z_{nm}; \rho, \beta)$ 며. $D(n, m)$ 값이 작을수록 영상의 형태에 대한 기여도가 높다. 트레이드마크 데이터베이스는 의사 저나이크 모멘트의 크기가 감마 분포에 의해 모델링 된다[2]. 그래서 $F(z_{nm}; \alpha, \beta)$ 는 감마 분포의 누적 밀도 함수(Cumulative Distribution Function)에 해당된다. 의사 저나이크 모멘트의 크기를 랜덤 변수로 하는 확률밀도함수(Probability Density Function)는 다음과 같다.

$$f(Z_{nm}; \alpha, \beta) = \frac{1}{\beta^\alpha} Z_{nm}^{\alpha-1} \exp(-Z_{nm}) U(Z_{nm})$$

- 742 -

$$\text{단, } \Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$$

그러나, 데이터베이스의 갯수와 영상의 종류에 따라 의사 저나이크 모멘트의 통계적인 특성이 바뀌는 경향이 있다. 그러므로 영상내의 물체의 형태에 대한 기여도가 높은 $D(n,m)$ 을 추출하기 위해 데이터베이스의 의사 저나이크 모멘트를 직접 이용하는 것이 일반적이다. 질의 영상의 n,m 차의 의사 저나이크 모멘트의 크기를 데이터 베이스 상의 동일한 차수의 의사 저나이크 모멘트 크기 를 비교하여 전체 데이터베이스에서 크기 우선 순위를 이용하여 $D(n,m)$ 를 정의할 수 있다

$$D(n, m) = \frac{\text{Rank}(Z_{nm})}{\text{데이터베이스총갯수}}$$

$D(n,m)$ 가 가장 적은 차수에 해당하는 n,m 차의 의사 저나이크 모멘트의 크기를 질의 영상의 대표 저나이크 모멘트로 정한다. 영상검색을 수행하기 위해 질의 영상의 대표 저나이크 모멘트의 차수에 해당하는 데이터베이스의 n,m 차 의사 저나이크 모멘트 크기와의 절대 거리에 대해서 유사한 영상 또는 동일한 영상을 검색할 수 있다.

3. 지역 형태 정보를 이용한 영상 검색 방법

영상에서 인터레스트 포인트를 추출하고 그 위치에서 미분 불변치 벡터와 위치 특성 벡터로 구성되는 특징 벡터를 계산하여 영상 검색을 수행하는 방식으로 이를 향상된 미분 불변치(IDI) 방식으로 정의하였다.

3.1 인터레스트 포인트 추출

인터레스트 포인트(Interest Points:IP)는 영상에서 신호가 2차원적으로 변하는 위치이며, 코너, T-Junctions, 텍스쳐가 급격히 변하는 위치에 해당한다. 현존하는 검출기 중에서 반복성(Repeatability)과 정보량(Information Content) 측면에서 Harris 검출기가 가장 우수함이 판명되었다. Harris 검출기는 자기상관함수와 관련된 매트릭스를 구하고, 자기상관함수의 중요 곡률(Curvatures)이 되는 이 매트릭스의 고유값(Eigenvalues)을 척도로 인터레스트 포인트를 검출한다[3]. 작은 이동에서 영상세기의 변화는

$$E(xy) = Ax^2 + 2Cxy + By^2 = (x, y) M(x, y)^T$$

단, $A = X^2 \otimes w B = Y^2 \otimes w C = (XY) \otimes w$

여기서, $M = \begin{bmatrix} A & C \\ C & B \end{bmatrix}$ 2 * 2 대칭 매트릭스

$$W_{u,v} = \exp(-(u^2 + v^2)/2\sigma^2) X = I \otimes (-1, 0, 1) = \partial I / \partial x$$

$$Y = I \otimes (-1, 0, 1)^T = \partial I / \partial y$$

M 은 회전에 불변하며, x, y 의 변화시 원점에서 분포상태의 척도가 된다. M 의 고유값 α, β 는 자기상관함수의 중요 곡률에 비례하므로 이를 기준으로 영상에서의 영역을 분류한다. 두개의 고유값이 모두 작으면 원도우 영역은 거의 동일한 픽셀값을 갖고, 고유값의 하나는 크고 하나는 작으면 원도우 영역은 에지성분에 해당되고, 고유값이 모두 크면 원도우 영역은 IP영역에 해당된다.

자기상관함수 매트릭스의 고유값은 다음과 같은 성질을 가지므로 고유값을 직접 구하지 않고 자기상관함수 매트릭스 성분을 바로 이용하여 실험에 필요한 IP영역을 판별한다.

$$Tr(M) = \alpha + \beta = A + B$$

$$Det(M) = \alpha\beta = AB - C^2$$

$$R = Det - kT^2 \quad k \text{는 } 0.25 \text{이하}$$

응답함수 R 은 A, B, C 의 합수이며 R 값을 척도로 R 값이 양수면 IP영역, 음수면 에지영역, 작으면 거의 동일한 픽셀값을 갖는다.

3.2 특징 벡터 계산

3.2.1 미분 불변치 벡터 생성

영상의 IP를 중심으로 하는 지역 원도우 영상은 미분

불변치 벡터로 나타낼 수 있다. 미분 불변치 벡터는 로컬젯(Local Jet)이라는 미분치 집합에 의해서 표현된다 [4]. 지역 원도우 영상을 I 라고 하고, 주어진 크기가 σ 일 때, 임의의 한점 $\mathbf{x} = (x_1, x_2)$ 에서 N 차 로컬젯은

$$J^N[I](\mathbf{x}, \sigma) = \{ L_{i_1, i_2, \dots, i_N}(\mathbf{x}, \sigma) | (\mathbf{x}, \sigma) \in I * R^+; n = 0, 1, 2, \dots, N \}$$

I 은 지역 원도우 영상(I)과 가우시안 함수 $G(\mathbf{x}, \sigma)$ 를 컨벌루션(convolution)한 성분이다. 첨자는 해당변수의 편미분에 해당된다. 컨벌루션과 미분연산의 교환법칙이 성립하여 다음과 같은 관계를 갖는다.

$$L_{i_1, i_2, \dots, i_N}(\mathbf{x}, \sigma) = I * \partial_i G = \partial_i I * G$$

로컬젯을 구하는 세 가지 방식 중에서 가우시안 함수를 편미분하여 지역 원도우 영상과 컨벌루션을 하는 방식이 가장 쉽고 안정적이다. 미분 불변치 벡터 DI 를 Einstein의 합방식과 직교좌표계 방식으로 나타내면 다음과 같다 [4][5].

$$DI[0, \dots, 4] = \begin{bmatrix} L \\ L_i L_i \\ L_i L_{ij} L_j \\ L_{ii} \\ L_{ij} L_{ji} \end{bmatrix} = \begin{bmatrix} L \\ L_x L_x + L_y L_y \\ L_{xx} L_x + 2L_{xy} L_x L_y + L_{yy} L_y L_y \\ L_{xx} + L_{yy} \\ L_{xx} L_{yy} + 2L_{xy} L_{yx} + L_{yy} L_{yy} \end{bmatrix}$$

$$DI[5, \dots, 8] = \begin{bmatrix} \varepsilon_{ij} (L_{jkl} L_i L_k L_l - L_{jlk} L_i L_l L_k) \\ (L_{iij} L_j L_k L_l - L_{ijk} L_i L_j L_k) \\ - \varepsilon_{ij} L_{jkl} L_i L_k L_l \\ L_{ijk} L_i L_j L_k \end{bmatrix}$$

(단, $\varepsilon_{12} = -\varepsilon_{21}$, $\varepsilon_{11} = -\varepsilon_{22} = 0$)

Einstein 합방식에서 첨자 i 는 변수들에 대한 편미분의 합을 나타낸다.

$$L_i = \sum_j L_{ij} = L_x + L_y, L_{ii} = \sum_j \sum_l L_{ijl} = L_{xx} + L_{yy} + L_{xy} + L_{yx}$$

미분 불변치 벡터는 영상크기 변화가 20%이상인 경우에는 멀티스케일 기법을 사용해야 한다. 멀티스케일기법은 질의 영상에 대하여 여러 크기의 가우시안 함수를 영상에 컨벌루션하여, 해당하는 미분 불변치 벡터를 각각 구하여 영상 검색에 사용한다. 영상의 크기변화가 2배까지 적용할 수 있기 위해서는 가우시안 함수의 크기는 0.4와 2.5사이에서 변하는 값에 해당하는 미분치 불변치 벡터를 구해야 한다[6].

3.2.2 위치특성 벡터 생성

위치특성 벡터는 영상에서 이미 추출된 IP의 위치 특성을 고려한 것이다. 그럼 1과 같이 중심으로부터 가장 가까운 다른 점들의 위치를 거리와 각도로 표현할 수 있다. 위치특성벡터는 다음과 같다.

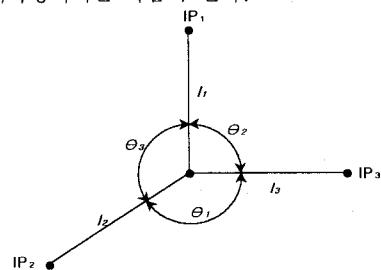


그림 1. 위치 특성 벡터

$$PC[0] = \min(l_1, l_2, l_3) \quad PC[3] = \min(\theta_1, \theta_2, \theta_3)$$

$$PC[1] = \text{median}(l_1, l_2, l_3) \quad PC[4] = \text{median}(\theta_1, \theta_2, \theta_3)$$

$$PC[2] = \max(l_1, l_2, l_3) \quad PC[5] = \max(\theta_1, \theta_2, \theta_3)$$

위치특성 벡터에서 길이에 해당하는 성분은 가장 적은 값부터 큰 값으로 정하고, 각도 성분 역시 적은 값부터 큰 값순으로 정한다. 위치특성 벡터가 크기변화에 불변하기 위해서 PC 벡터 성분 중에서 길이에 해당하는 성분은 정규화해야 한다.

4. 실험 결과

4.1 실험

본 연구에서는 128×128 크기인 300개의 트레이드마크 영상 데이터베이스를 이용하여 영상 검색에 대한 실험을 수행하였다. 기존의 방식인 의사 저나이크 모멘트(ZM) 방식과 미분 불변치(DI) 방식을 각각 실험하고 이의 특성을 각각 향상시킨 IZM 방식과 IDI 방식으로 각각 실험을 하고, 이 두가지 방식을 복합한 IZM&IDI 방식으로 실험하여 성능을 비교한다.

4.2 실험 결과

4.2.1 영상의 변형에 따른 동일한 영상을 찾는 실험

표 1, 2, 3은 질의 영상에 대해 영상 내에서의 객체의 기하학적 변형(회전, 이동, 축척), 영상 열화(잡음첨가), 조명 변화되었을 때 영상 검색을 수행한 결과이다. 실험치는 데이터베이스의 검출 순위를 상대적인 검출률로 환산한 것이다. 실험치(%)는 ($(\text{데이터베이스 총 갯수} - \text{검출 순위}) / \text{데이터베이스 총 갯수} * 100$)으로 나타내었다.

표 1. 영상의 이동, 회전, 축척에 따른 영상 검색 결과

영상의 기하학적 변형	ZM	DI	제안1 (IZM)	제안2 (IDI)	제안3 (IZM & IDI)
이동	100%	100%	100%	100%	100%
회전10	89.7%	99.0%	93.0%	95.7%	99.3%
회전45	97.9%	98.7%	99.0%	97.3%	100%
회전90	100%	98.7%	100%	97.3%	100%
크기90%	93.6%	100%	97.7%	99.0%	100%
크기110%	94.7%	100%	93.0%	93.7%	98.7%

4.2.2 질의 영상에 대한 유사성 비교 검색 결과

질의 영상이 그림 2의 (a)와 같을 때 데이터베이스 상에서 유사한 영상을 찾는 실험을 기존의 방식과 제안한 방식으로 영상 검색을 수행한 결과이다.



(a) 질의 영상

Rank	ZM	DI	IZM & IDI
------	----	----	-----------



(b) 검색 결과

그림 2. 내용 기반 영상 검색 결과

표 2. 영상에 잡음이 첨가되었을 때 영상 검색 결과

염화된 영상의 PSNR	ZM	DI	제안1 (IZM)	제안2 (IDI)	제안3 (IZM & IDI)
28dB	100%	100%	100%	100%	100%
26dB	99.0%	100%	100%	99.3%	100%
24dB	94.3%	100%	99.0%	99.3%	100%
22dB	88.7%	99.3%	97.7%	99.3%	100%
20dB	66.7%	100%	90.0%	96.3%	98.7%

표 3. 조명이 변화되었을 때 영상 검색 결과

조명변화	ZM	DI	제안1 (IZM)	제안2 (IDI)	제안3 (IZM & IDI)
-50	100%	31.7%	100%	55.6%	92.7%
-20	100%	55.3%	100%	63.6%	92.3%
-10	100%	56.0%	100%	67.0%	91.7%
+10	100%	98.7%	100%	99.3%	100%
+20	100%	80.7%	100%	99.3%	100%
+50	100%	80.0%	100%	99.3%	100%

4. 결론

본 논문에서는 형태 정보를 이용하여 질의 영상을 데이터베이스에서 검색하는 방법으로 IZM 방식, IDI 방식, IZM&IDI 방식을 제안하였다.

IZM 방식은 의사 저나이크 모멘트 중에서 영상 전체의 형태 기여도가 가장 큰 모멘트만 고려함으로써 기존의 ZM 방식에서의 계산량을 줄였고, 영상의 기하학적 변형과 영상 열화 경우에도 좀더 강인한 결과를 보였다. IDI 방식은 기존의 DI 방식에 위치특성 벡터를 추가함으로써 영상의 분별성을 높이고 조명 변화에 좀더 강인한 결과를 얻었다. 두 방식을 복합한 IZM&IDI 방식은 기하학적 변형, 조명 변화, 영상 열화에 모두 강인한 결과를 보였다.

질의 영상의 유사성 비교 실험에서 ZM 방식은 영상 내에 포함된 객체의 내부 피셀 값은 다소 무시되고 객체의 전체 형태가 유사한 순서로 검출되는 반면, DI 방식은 객체의 전체 형태는 다소 무시되고 객체의 내부 피셀 값이 유사한 영상이 우선 검출되었다. IZM&IDI 방식은 영상내에 객체의 전체 형태와 지역 형태 및 내부의 피셀 값이 동시에 고려되어 데이터베이스에서 형태가 가장 유사한 영상을 검색함을 보였다.

(참고문헌)

- (1) A. Khotanzad and Y.H. Hong, "Invariant Image Recognition by Zernike Moments," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.12, pp.489-498, 1990.
- (2) Yong-Sung Kim and Whoi-Yul Kim, "Content-Based Trademark Retrieval System Using A Visually Salient Feature" Image & V computing, vol.16, no.12-13, 931-939, 1998.
- (3) C. Harris and M. Stephens, "Combined Corner and Edge Detector," Alvey Vision Conference, pp.147-151, 1988.
- (4) J.J. Koenderink, and A.J. van Doorn, "Representation of Local Geometry in the Visual System," Biological Cybernetics, vol.55, pp. 367-375, 1987.
- (5) Cordelia Schmid and Roger Mohr, "Local Grayvalue Invariants for Image Retrieval," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.19, no.5, pp.530-535, 1997.
- (6) Cordelia Schmid and Roger Mohr, "Matching by Local Invariants," Technical report N2644, INRIA, Aug,1995.