

KLT 알고리즘을 사용한 MPEG-7의 최적Descriptor 조합선택에 관한 연구

김 현 민, 권 기 상, 권 혁 민, 최 윤 식
연세대학교 전기전자학과
전화 : 02-2123-2774 / 핸드폰 : 018-335-4785

Selecting a Composite of MPEG-7 Visual Descriptor by KLT Algorithm

Hyun-Min Kim, Ki-Sang Kwon, Hyuk-Min Kwon, Yoon-sik Choe
Dept. of Electrical Eng, Yonsei University
E-mail : sleaky76@hanmail.net

Abstract

Due to the increasing of multimedia data quantity, database searching based on image becomes important. For this scheme, MPEG-7 gives a good solution to efficient data searching. MPEG-7 uses Descriptors which are color, texture, and shape to extract features from images. It is obvious that using more than one Descriptor causes more accurate data searching result than using just one. In this paper, selecting a composite of MPEG-7 visual Descriptor using KL-Expansion is proposed.

I. 서 론

멀티미디어 데이터의 증가로 인해 이미지 기반의 검색 방법이 점차 중요해 지고 있다. 단순한 텍스트 기반의 검색은 사람의 주관적인 판단에 기초하여 수동적인 메커니즘을 사용하기 때문에, 멀티미디어 데이터 검색의 방법으로는 부 적절하다. 이미지 기반 검색의 한 가지 솔루션으로서 MPEG-7은 효과적인 descriptor를 통해 영상의 feature를 추출해 낸다. MPEG-7에서는 여러 descriptor들이 정의되어 있는데, visual descriptor에는 크게 low-level 정보로서 color descriptor, texture descriptor, shape descriptor 등이

있다. 이러한 descriptor들은 서로 상이한 알고리즘으로 영상의 feature들을 뽑아낸다. Color descriptor는 색의 분포에 대한 기준으로 feature를 뽑아내고, texture descriptor는 영상의 질감에 대한 feature를 뽑아낸다. 또한 shape descriptor는 영상의 모양에 대한 정보를 뽑아낸다. 이와 같이 여러 descriptor들의 feature extraction method가 다르므로 서로 상호 보완적인 descriptor를 두 가지 이상 사용한다면, 영상 검색의 결과가 좋아 질 것임을 쉽게 예측할 수 있다.

본 논문에서는 두 가지 이상의 descriptor들을 효과적으로 묶어서 description scheme을 만들 때 어떠한 descriptor들의 조합으로 scheme을 구성하는 것이 가장 효과적일지를 정하였다. 이를 위해 KLT[1][2] 변환을 사용하여 여러 descriptor들 중에 가장 효과적인 descriptor조합을 결정하였다.

II. MPEG-7 Descriptor

MPEG-7[3] 엔진은 이미지 기반의 분류-검색 알고리즘을 사용함으로써, 영상들의 패턴인식과 그 결과로서의 효과적인 데이터베이스의 구축을 가능케 한다. MPEG-7 엔진은 영상을 분석하여, 여러 가지 의미 있는 feature들을 영상으로부터 뽑아낸다. 이때 feature extraction을 수행하기 위해 descriptor를 사용하게 되

는 데, 현재 MPEG-7내에는 여러 가지 descriptor들이 정의되어 있다. 이렇게 얻어진 feature에 대한 정보를 여러 개 조합한다면 더 좋은 결과를 나타낼 수 있고 이때 feature 정보들을 묶는 역할을 하는 것이 description scheme 이다. 본 논문에서는 두 가지 이상의 descriptor들을 효과적으로 묶어서 description scheme을 만들 때 어떠한 descriptor를 사용할 지에 대한 결정을 내리기 위한 잣대로서 KL-Expansion 방법을 이용하였다.

MPEG-7 visual standard의 주된 목표는 streamed되거나 저장된 이미지의 표준화된 descriptor를 제공하는 것이다. Descriptor는 기술하는 방법에 따라 크게 두 가지 부류로 나눌 수 있는데, 하나는 low-level descriptor 이고 다른 하나는 high-level descriptor이다. 이때, low-level descriptor는 영상의 색이나 질감, 모양 등에 대한 정보로서 완전히 자동으로 추출될 수 있으며 color (색상), texture (질감), shape (형태), motion (운동), localization 등으로 구성되어있다. 반면, high-level descriptor는 추상 단계기술에 관한 내용으로, 의미론적 정보를 담고 있다.

III. KL-Expansion의 적용

3.1 KL-Expansion 적용의 이유

특정 선택의 문제에 있어서 여러 가지의 접근 방법이 가능한데, 이에에는 최소 엔트로피 접근 방법과 직교전개 접근 방법[4] 등이 있다. 최소 엔트로피 접근 방법은 패턴 부류가 정규 분포를 따를 경우 적용 가능하고, 직교전개 접근 방법은 그렇지 않은 경우에 해결방법을 제공해 준다. 본 논문에서는 MPEG-7 descriptor의 계수들의 확률적 분포가 정규 분포를 따르지 않는다는 전제하에 매트릭스의 차원을 줄이는 데에 KL-Expansion 방법을 적용하였다.

3.2 변수의 정의

KL-Expansion을 적용하기 전에 다음과 같이 변수를 설정하였다.

M: Descriptor($w_1, w_2 \dots w_M$) 수의 개수

$x_i(t)$: M개의 descriptor중 임의의 한 descriptor로부터의 관찰 값

C_{ij} : $E(C_{ij})=0$ 을 만족하는 확률계수

$\Phi_j(t)$: Deterministic orthogonal 함수들의 집합

$x_i(t)$ 가 임의의 영상에서 어느 특정한 descriptor를 사용해 균일하게 표본 추출되었다면,

$$X_i = [x_i(t_1), x_i(t_2), \dots, x_i(t_3)]^T$$

단, (n: $x_i(t)$ 의 표본수) (1)

와같이 표현되어지고, 동시에

$$X_i = \sum_{j=1}^n C_{ij} \Phi_j$$
 (2)

와같은 표기도 가능하다. 이때, $E(C_{ij})=0$ 을 가정 하였고,

$$\Phi_j = [\Phi_j(t_1), \Phi_j(t_2), \dots, \Phi_j(t_3)]$$

$$C_i = [C_{i1}, C_{i2}, \dots, C_{in}]$$

로 정의하였다. 그렇다면 식(2)은 다음과 같이 나타낼 수 있다.

$$X_i = \sum_{j=1}^n C_{ij} \Phi_j = \Phi C_i$$
 (3)

단, ($\Phi = \{ \Phi_1, \Phi_2, \dots, \Phi_n \}$)

3.3 Eigenvalue 의 도출

자기상관 행렬은

$$R = \sum_{i=1}^M P(w_i) E\{X_i X_i^T\}$$
 (4)

로 표시되고 식(3)를 이용하면 자기상관 행렬은

$$R = \sum_{i=1}^M P(w_i) E\{\phi C_i C_i \phi^T\} = \Phi \sum_{i=1}^M P(w_i) E\{C_i C_i^T\} \Phi^T$$
 (5)

와같이 된다. 왜냐하면 Φ 는 deterministic이기 때문이

다. 이제,

$$\sum_{i=1}^M P(w_i) E\{C_i C_i^T\} = D_\lambda$$

$$D_\lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

를 가정하면 식(5)은

$$R = \Phi D_\lambda \Phi^T \tag{6}$$

가 된다. 기저벡터 Φ_j 가 orthonormal 이라고 가정하였으므로, $\Phi \Phi^T = I$ 가 된다. 따라서, 식(6)는 $R\Phi = \Phi D_\lambda \Phi^T \Phi = \Phi D_\lambda$ 로 변형 가능하다. 즉,

$$R\Phi_j = \lambda_j \Phi_j \tag{7}$$

이다. 식(7)에서 좌변 보다, 우변의 차원이 한 차원 낮음에서부터 알 수 있듯이 λ_j 는 eigenvalue가 되고, Φ_j 는 eigenvector가 된다.

3.3 Eigenvalue의 선택에 의한 Descriptor의 선택

Descriptor 종류 w_i 의 bin X_i 에 있어서 $C_i = \Phi^T X_i$ 의 관계가 성립한다. 왜냐하면 (식3)에서 $\Phi C_i = X_i$, 이고 이 식은 $\Phi^T \Phi C_i = \Phi^T X_i$ 을 거쳐, $C_i = \Phi^T X_i$ 로 정리되기 때문이다. 그런데 자기상관 행렬 R의 eigenvalue들중 큰값 m ($m < n$)개를 선택한다면, Φ^T 의 크기는 $m \times n$ 이므로 C_i 는 X_i 보다 더 낮은 차원의 상 벡터임을 알 수 있다. 따라서 행렬 A를 다음과 같이 표기하는 것이 가능하다.

$$A = \Phi^T = \begin{bmatrix} \Phi_1^T \\ \Phi_2^T \\ \vdots \\ \Phi_m^T \end{bmatrix} \tag{8}$$

본 논문에서는 edgehistogram, homogeneous texture, colorlayout, dominantcolor descriptor를 사용하였고, 따라서 $M=4$ 가 된다. 이 네 가지 descriptor를 임의의 한 영상에 적용하였을 경우, 각 descriptor마다 서로 다른 수의 계수들을 가지므로, 각 descriptor에 해당하는 eigenvalue를 더한 값을 해당 descriptor를 대표하는 eigenvalue set으로 정한다. 이렇게 정한 4개의 eigenvalue set중 가장 큰 값을 갖는 set을 해당 영상을 가장 잘 표현하는 descriptor로 정하였다. 본 연구에서는 가장 효과적인 descriptor의 조합을 고르는 데 목적이 있으므로, 큰 값을 가지는 eigenvalue set 두개를 선택하여 descriptor 조합으로 결정하였다.

IV. 실험 결과

본 논문에서는 edghistogram, homogeneous texture, colorlayout, dominantcolor, 이상의 4개의 descriptor들중 임의의 특정한 영상을 표현하는데 가장 효과적인 두 가지 descriptor를 선택하는 기준으로 descriptor의 계수 값들의 자기 상관 행렬의 eigenvalue를 사용하였다. Descriptor들의 계수를 랜덤 변수로 사용하기 위해서는 각 계수들의 standardization이 선행되어야 한다. standardization을 함으로써, 각 랜덤 변수들은 평균이 0, 분산이 1인 정규 분포를 따르게 된다. 이 경우, 각 랜덤 변수들의 평균과 분산이 다르기 때문에 standardization은 각각 따로 수행해야한다. 모든 랜덤 변수의 값의 범위는 MPEG-7 standard의 내용을 따랐으며, colorlayout의 경우는 YDCCoeff, CbDCCoeff, CrDCCoeff, YACCoeff, CbACCoeff, CrACCoeff 계수의 값의 범위가 정해져 있지 않으므로, 임의의 영상 100개를 조사했을 때의 최대, 최소 값의 범위로 근사 화 하였다.

표1. eigenvalue의 분포

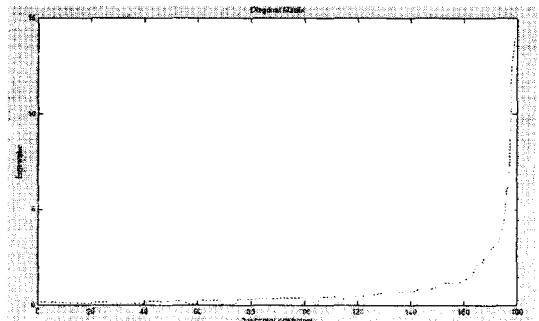


표1. 에서 x축에는 descriptor의 종류별로, 계수들을 나열하였다. 처음 25개의 eigenvalue 값들은 dominantcolor에 해당하는 값이고, 그 다음 12개는 colorlayout에, 다음 80개는 edgehistogram에, 마지막 62개는 homogeneoustexture에 해당하는 값들이다. dominantcolor 에 해당하는 eigenvalue set의 값은 3.36이고, colorlayout에 해당하는 eigenvalue set 값은 1.35, edgehistogram에 있어서는 19.26, homogeneoustexture에 있어서는 110.45의 값을 얻을 수 있었다. 따라서, homogeneoustexture와 edgehistogram을 선택하여 조합을 만드는 것이 가장 서로 보완적이며, 검색 효과가 우수하다고 할 수 있다.

V. 결론

KL-Expansion 방법을 통한 descriptor의 선택을 실험적 방법을 통한 Descriptor의 선택 결과와 비교해보면, 두 가지 결과의 순서가 유사 함을 알 수 있다. KL-Expansion을 통해 최적 descriptor로 결정된 descriptor는 homogeneoustexture와 edgehistogram였는데, 실험적 결과에 토대한 최적 descriptor 결과와 일치하였다. 이때, 실험적 방법으로는 XM을 사용하여 100개의 영상 데이터베이스에서 쿼리 영상과 유사한 10개의 영상을 검색 하였을때 10개의 영상내에 쿼리 영상과 유사한 영상이 어떠한 순위를 가지는 가로 판단하였다.

아래에는 실험적 방법을 수행할 때, 쿼리 영상과 쿼리영상과 유사하다고 판단될 수 있는 영상, 그리고, 같은 데이터 베이스 내에 존재하지만 쿼리 영상과 유사하다고 판단 될 수 없는 영상을 나타내었다.



그림1. 쿼리 영상



그림2. 유사영상



그림3. 쿼리영상과 유사하지 않은 영상

아래의 표2.는 XM을 사용하여 분류를 하였을 때 전체 100개의 영상으로 이루어진 database에서 쿼리 영상과 유사한 영상을 찾았을 경우 유사하다고 판단된 최상위 10개 영상을 표현한다. 표에서 알 수있듯이 homogeneoustexture와 edgehistogram이 쿼리영상을 가장 잘 표현함을 알 수있다.

표2. XM을 사용한 검색 결과

DC	CL	EH	HT
○	○	○	○
○	○	○	○
○	△	○	○
△	○	△	○
○	○	○	△
△	×	○	○
○	△	×	×
×	○	△	△
×	×	×	×
×	×	×	×

DC: dominantcolor CL: colorlayout
 EH: edgehistogram HT: homogeneoustexture
 ○ : 올바르게 검색된 결과
 △ : 검색결과 불분명
 × : 잘못된 검색결과

참고문헌

[1] Richard O.Duda, Peter E.Hart, and David G. Stork, *Pattern Classification*, John Wiley & Sons, New York.
 [2] S. Watanabe, "Karhunen-Loeve Expansion and Factor Analysis-Theoretical Remarks and Applications," *Proc. of the 4th Conference on Information Theory*, Prague, 1965.
 [3] B. S. Manjunath, Philippe Salembier, and Thomas Sikora, *Introduction to MPEG-7*, Wiley.
 [4] Andrew Webb, *Statistical Pattern Recognition*, Wiley.