

키 프레임 특징들에 적응적 가중치 부여를 이용한 검색 성능 개선

김 강 욱[†]

요 약

비디오 검색 및 색인은 먼저 압축 비디오에서 장면전환을 검출하여 샷(shot)으로 분리한 후 샷 내에 키 프레임 특징 정보들의 유사도 비교를 통해 이루어진다. 일반적으로 내용기반 영상 및 비디오 검색에서는 컬러, 형태, 질감의 세 가지 대표적인 영상 특징들이 주로 사용된다. 그러나 여러 특징들이 결합되어 사용되는 검색 시스템이라 할지라도 각 특징들에 대한 가중치가 적합하게 부여되지 않으면 검색되는 결과 영상의 순위가 크게 변하여 검색 성능이 떨어지게 된다. 이러한 문제점을 해결하기 위해 본 논문에서는 여러 특징들이 결합되어 사용될 때 각 특징에 대한 가중치를 적응적으로 부여해서 비디오 검색 성능을 개선하고자 한다. 제안한 방법을 3,200개 키 프레임으로 구성된 비디오 데이터베이스에서 실험을 하였고 다양한 성능평가 방법을 통해 제안한 방법이 기존 고정가중치 부여를 이용한 방법과 비교하여 검색 성능이 개선됨을 볼 수 있었다.

Improvement of Retrieval Performance Using Adaptive Weighting of Key Frame Features

Kang-Wook Kim[†]

ABSTRACT

Video retrieval and indexing are performed by comparing feature similarities between key frames in shot after detecting a scene change and extracting key frames from the shot. Typical image features such as color, shape, and texture are used in content-based video and image retrieval. Many approaches for integrating these features have been studied. However, the issue of these approaches is how to appropriately assign weighting of key frame features at query time. Therefore, we propose a new video retrieval method using adaptively weighted image features. We performed computer simulations in test databases which consist of various kinds of key frames. The experimental results show that the proposed method has better performance than previous works in respect to several performance evaluations such as precision vs. recall, retrieval efficiency, and ranking measure.

Key words: Key Frame(키 프레임), Retrieval(검색), Video(비디오), Weighting(가중치)

1. 서 론

최근 멀티미디어 기술의 발달과 다양한 하드웨어의 발전에 힘입어 여러 가지 형태의 데이터들이 기하급수적으로 생겨나고 있을 뿐만 아니라 디지털화됨

으로 인해서 멀티미디어 정보에 대한 이용과 가공이 용이해 지고 있다. 하지만, 대용량화에 따른 효율적인 데이터의 저장 및 관리가 필요하게 되었으며, 더불어 사용자의 요구를 만족시킬 수 있는 데이터의 검색이나 색인이 중요한 문제로 대두되고 있다. 특

※ 교신저자(Corresponding Author) : 김강욱, 주소 : 경기도 화성시 석우동 예당마을 롯데캐슬 148동 901호(445-855), 전화 : 010-2678-2501, E-mail : ekans999@gmail.com

접수일 : 2013년 10월 22일, 수정일 : 2013년 11월 20일
완료일 : 2013년 11월 25일

[†] 삼성전자

히, 최근에는 뮤직비디오나 영화, TV 프로그램과 같은 동영상 정보들이 MPEG과 같은 압축 기술에 의해 VOD 형태로 저장 및 제공되고 있다. 이러한 정보들을 분류 및 관리하거나 사용자의 이용을 용이하게 하기 위한 비디오 검색 및 구조화 방안들이 많이 제안되어 왔다[1-2]. 특히 비디오는 멀티미디어 정보량의 대부분을 차지할 뿐만 아니라 교육이나 오락과 같은 정보전달을 위한 일반적인 수단으로 사용되고 있다. 그래서 비디오 데이터베이스 구축 및 검색을 위해서는 한 편의 비디오를 비디오 분류의 기본 단위인 샷(shot)으로 분할한 후 샷 내의 키 프레임 특징들을 추출하여 이를 기반으로 검색하고 분류하는 것이 반드시 필요하다.

일반적인 내용기반 영상 및 비디오 검색 방법에는 크게 컬러, 형태, 질감과 같은 세 가지 대표적인 영상의 특징들을 이용하게 된다. 컬러를 이용한 검색 방법은 영상에 대해 히스토그램을 비교하는 방법으로 초기 내용 기반 검색에서 많이 사용되었다. 형태적 특징을 이용한 방법은 특히 개체가 뚜렷이 구별되는 영상, 즉 상표나 사람의 몸짓, 얼굴모양 등의 검색에 주로 적용되고 있다. 컬러, 형태 정보 외에 물체 표면의 특성을 나타내는 질감 정보가 검색에 이용될 수 있다. 그러나 한 가지 특징만을 중점적으로 사용하는 검색 방법은 키 프레임의 내용이 복잡해지고 비교해야 할 프레임의 수가 많아질수록 좋은 검색 성능을 내지 못한다. 그래서 이러한 단점을 보완하고 보다 다양한 비디오 데이터베이스에 적용하기 위해 여러 가지 영상 특징들을 결합하는 방법들이 제안되었다 [3-4]. 그러나 여러 특징들을 결합해서 사용하는 검색 시스템이라 할지라도 일반 사용자들이 각 특징들에 대한 가중치를 적합하게 결정하기 힘들다. 더구나 키 프레임의 특징들은 서로 다른 통계적인 특성을 보이므로 그들 간에 가중치가 올바르게 부여되지 않으면 유사 프레임의 검색 순위가 크게 변하여 검색 성능이 저하되게 된다. 이러한 문제점을 해결하기 위해 본 논문에서는 컬러, 형태, 질감에 대한 전처리 과정을 이용하여 키 프레임 영상을 개략적으로 구별하고 여기서 얻어진 정보를 데이터베이스로 미리 구축한 뒤, 질의 영상과 데이터베이스 내의 키 프레임을 비교할 때 각 특징들에 대한 가중치를 적응적으로 부여하여 사용자 편의성 제공 및 검색 성능을 향상시킨 새로운 방법을 제안한다.

본 논문의 구성은 2장에서 장면전환 검출 및 키 프레임 추출에 대해서, 3장에서는 대표적인 영상 특징인 컬러, 형태, 질감을 표현하는 방법을 살펴보고 4장에서는 적응적 가중치 부여를 위해 필요한 전처리 과정과 가중치 결정 방법을 설명한다. 그리고 5장에서는 다양한 성능 평가 측면에서 실험을 하여 제안한 방법의 타당성을 검증하고 6장에서 결론을 맺는다.

2. 장면전환 검출 및 키 프레임 추출

MPEG 비디오에서 장면전환 검출을 위한 다른 중요한 파라미터로는 세 가지 매크로블록 타입에 관한 정보들이며 P-프레임과 그것의 참조 프레임간의 관계 혹은 B-프레임과 참조 프레임간의 관계를 나타낸다. FMB(forward macroblock), BMB(backward macroblock), BIMB(bidirectional macroblock)의 수는 현재 프레임과 그것의 과거, 미래 또는 과거와 미래 프레임들에 얼마나 가까운가를 나타낸다. 이런 매크로블록의 수가 많아질수록 현재 프레임이 참조 프레임과 더 가깝게 된다. 최근에 MPEG으로 압축된 데이터의 부분 복호화에 근거한 여러 가지 장면전환 검출 알고리즘이 제안되고 있으며 사용되는 정보에 따라 크게 세 가지로 분류된다. DCT 계수에 근거한 알고리즘, 움직임 벡터에 근거한 알고리즘, 둘 모두를 이용한 방법이 있다. 여러 가지 다양한 장면전환 검출 기법이 있지만 본 논문에서는 비교적 간단하며 속도가 빠를 뿐 아니라 장면전환의 정확한 위치까지 검출할 수 있는 방법[5]을 사용하였다. 사용된 알고리즘의 상세 과정은 다음과 같다.

먼저 연속되는 GOP사이에 두 개의 I-프레임을 비교하기 위해서, 두 개 I-프레임 f_m 과 f_n 사이에 정규화된 DC 계수값 절대차의 평균을 정의한다.

$$D(f_m, f_n) = \frac{1}{K} \sum_{i=1}^K \frac{|c(f_m, i) - c(f_n, i)|}{\max(c(f_m, i), c(f_n, i))} \quad (1)$$

여기서 $c(f_p, i)$ 는 프레임 f_p 에서 블록 i 의 DC 계수값이며 K 는 I-프레임에서 블록의 개수이다. $D(f_m, f_n)$ 이 임계치 이상이 될 때, 현재 I-프레임 f_m 과 지난 I-프레임 f_n 사이에 장면전환이 일어났다고 판단한다. 좀더 정확한 위치를 알아내기 위해서 동일한 GOP 내의 P-프레임들에 대해서 식(2)의 T_p 값을 이용한다.

$$T_p = \frac{(\text{Number of intra MB}) - (\text{Number of FMB})}{(\text{The total number of MB})} \quad (2)$$

T_p 는 현재 P-프레임과 그의 참조 프레임(I or P-프레임)사이의 관계를 나타낸다. T_p 가 음이거나 0에 가까우면 장면전환이 일어나지 않았으며 T_p 가 1에 가까우면 장면전환이 일어났다고 판단한다. P-프레임의 경우와 마찬가지로 B-프레임을 조사하기 위해서 식(3)과 식(4)로 주어진 두 개의 비 T_b 와 T_f 를 사용한다.

$$T_b = \frac{(Number\ of\ BMB) + (Number\ of\ BLMB)}{(The\ total\ number\ of\ MB)} \quad (3)$$

$$T_f = \frac{(Number\ of\ FMB) + (Number\ of\ BLMB)}{(The\ total\ number\ of\ MB)} \quad (4)$$

T_b 는 현재 B-프레임과 그의 미래 참조 프레임사이의 관계를 T_f 는 현재 B-프레임과 그의 과거 참조 프레임사이의 관계를 나타낸다. T_f 와 T_b 가 0에 가까우면 현재 B-프레임에서 장면전환이 일어났다고 판단한다. 만약 두 개의 연속되는 B-프레임 모두에서 장면전환이 발생되었다면 첫 번째 B-프레임에서 장면전환이 일어났다고 판단한다. 그리고 T_b 가 0에 가깝고 T_f 가 0에 가까운 값이 아니면 미래 참조 프레임에서 장면전환이 일어났다고 판단한다. T_b 와 T_f 모두 0에 가까운 값이 아니면 과거 참조프레임과 미래 참조프레임 사이에 장면전환이 일어나지 않은 것이다. 장면전환 검출에 의해 분리된 샷에서 해당 키 프레임은 장면전환이 일어난 프레임 이후의 첫 번째 I-프레임으로 선정한다.

3. 키 프레임 특징 추출

3.1 컬러(color)

컬러는 입력되는 영상의 회전이나 이동과 같은 변화에 크게 영향을 받지 않으므로 영상 검색에 있어서 아주 중요한 특징중의 하나이다. 컬러를 나타내는 방법에는 RGB(Red, Green, Blue), HSI(Hue, Saturation, Intensity) 같은 컬러 공간을 사용한다. 컬러 히스토그램을 기반으로 하는 검색은 형태나 질감을 이용한 검색에 비해 속도가 빠르다. 3개의 1-D 히스토그램을 이용한 검색 방법에는 Histogram Intersection과 Histogram Distance 방법이 있다. 일반적으로 컬러 히스토그램 사이의 유사도를 측정하기에는 Histogram Distance 방법이 Histogram Intersection에 비해 더 좋은 성능을 갖는다고 알려져 있다[3]. 그래서 본 논문에서는 유사도 계산을 위해 Histogram

Distance 방법을 사용한다. Histogram Distance 방법에 사용되는 Euclidian Distance는 특징 벡터들 사이의 차이를 계산하는 방법이다. I_R, I_G, I_B 를 데이터 영상들의 정규화된 컬러 히스토그램이라 하고 Q_R, Q_G, Q_B 를 질의 영상의 정규화된 컬러 히스토그램이라 할 때 질의 영상과 데이터 영상들과의 Histogram Distance에 의한 유사도 $S_c^{ED}(I, Q)$ 는 식 (5)와 같다.

$$S_c^{ED}(I, Q) = 1.0 - \sqrt{\frac{\sum_r (I_R(r) - Q_R(r))^2 + \sum_g (I_G(g) - Q_G(g))^2 + \sum_b (I_B(b) - Q_B(b))^2}{2 \times 3}} \quad (5)$$

3.2 형태(shape)

형태적인 특징을 추출하는 방법에는 Chain Coded String, Fourier Descriptor, 에지 분포를 이용한 방법들이 있다. 본 논문에서는 에지 분포를 이용해서 영상 객체의 형태적 특징을 추출하고 이를 인덱스 키로 사용할 수 있는 수치로 나타내기 위해 에지 성분이 각 부 영역에서 차지하는 양을 특징 값으로 한다. 형태적 유사도 $S_s(I, Q)$ 는 질의 영상과 데이터 영상들 사이의 형태적 키 값들의 차를 합한 후 이를 전체 에지 화소의 개수로 나눈 값으로 정의한다. 이 때 부 영역 I 에서 데이터베이스 영상들의 에지 개수를 $I_s(i)$ 라 하고 질의 영상의 에지 개수를 $Q_s(i)$ 라 한다.

$$S_s(I, Q) = 1.0 - \frac{\sum_{i=1}^{256} |I_s(i) - Q_s(i)|}{\sum_{i=1}^{256} (I_s(i) + Q_s(i))} \quad (6)$$

3.3 질감(texture)

시각적으로 물체를 식별하기 위해서는 컬러, 형태 정보 외에 물체 표면의 특성을 나타내는 질감 정보가 이용될 수 있다. 질감정보의 표현 방법은 통계적 방법과 구조적 방법 두 가지가 있다. 본 논문에서는 통계적 질감 묘사방법인 co-occurrence 행렬을 사용한다. 질감에 대한 키 생성은 4 방향(0°, 45°, 90°, 135°)에 대한 256×256 co-occurrence 행렬을 구한 뒤 각각에 대해서 Energy, Entropy, Correlation, Local Homogeneity, Inertia 등 5가지 질감 특징을 구하여 모두 20개 키를 생성한다. 질감에 대한 유사도 계산을 위한 $I_t(i, j), Q_t(i, j)$ 는 데이터 영상과 질의 영상의 질감 특징을 나타낸다. i 는 각도를 나타내는 인덱스

로써 0°, 45°, 90°, 135°를 나타내고 j는 5개의 질감 특징인 E(energy), H(entropy), C(correlation), L(local homogeneity), I(inertia)를 나타낸다. 질감에 대한 유사도 $S_t(I, Q)$ 는 식(7)과 같다.

$$S_t(I, Q) = 1.0 - \frac{\sum_{i=1}^4 \sum_{j=1}^5 |I_t(i, j) - Q_t(i, j)|}{\sum_{i=1}^4 \sum_{j=1}^5 (I_t(i, j) + Q_t(i, j))} \quad (7)$$

4. 키 프레임 특징들에 적응적 가중치 부여를 이용한 검색

4.1 처리(preprocessing)

비디오에서 추출된 키 프레임 영상들은 그림 1과 같은 전처리 과정을 거쳐 그 결과를 간단히 비트(bit)로 나타내고 유사도 계산에 필요한 키 프레임 특징들을 수치화하여 그림 1과 같이 데이터베이스 내에 저장한다.

이 단계에서 사용되는 요소로써 컬러, 형태, 질감의 3가지 특징을 사용하여 키 프레임을 다음과 같이 구분하고 이를 3 비트로 간략히 표현한다.

- Color Image* → '1'
- Black and White Image* → '0'
- Image with Clear Object(s)* → '1'
- Image without Any Objects* → '0'

- Texture Image* → '1'
- Non-Texture Image* → '0'

컬러에 대한 전처리 과정에서는 컬러 영상과 흑백 영상을 구별하는 위해 R, G, B에 해당하는 히스토그램 $h(r)$, $h(g)$, $h(b)$ 를 구해서 그림 2와 같은 과정으로 구분한다.

영상에서 개체의 구분이 뚜렷한 지의 여부는 PIM (Picture Information Measure)[6] 분산에 따라 구별할 수 있으므로 형태에 관한 전처리 과정으로 PIM의 분산을 이용한다. 임계치는 실험을 통해 전체 데이터베이스 내에 있는 영상 개수의 반 정도를 포함하도록 결정하였다.

질감에 대한 전처리에서는 질감영상과 비 질감영상의 구분을 위해 영상을 웨이브릿 변환하여 Decomposition Level 1에 있는 고주파 영역의 에너지 크기를 이용한다. 질감 영상일수록 높은 에너지를 가지므로 임계치는 실험적인 값인 25,000을 사용하였다.

4.2 적응적 가중치 부여를 이용한 키 프레임 검색

제안한 검색 시스템을 이용한 실제 키 프레임 검색 과정은 그림 5와 같다.

질의 영상이 주어지면 데이터베이스를 만드는 과정과 동일한 전처리를 수행하고 프레임 특징을 수치화 한 후 데이터베이스내의 키 프레임들과 비교한다.

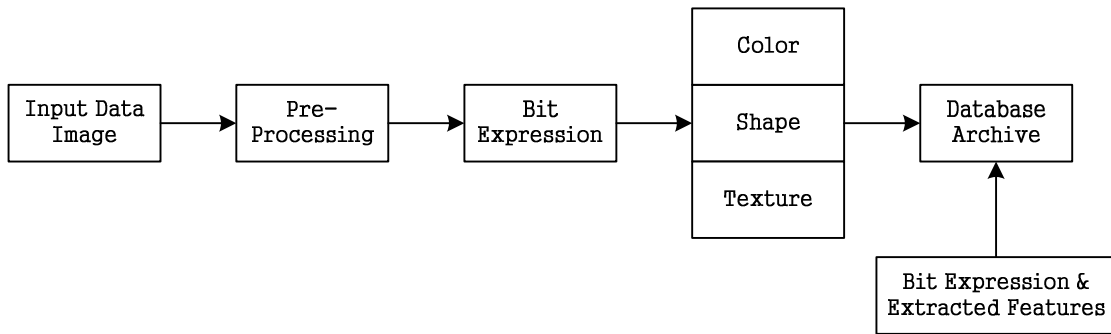
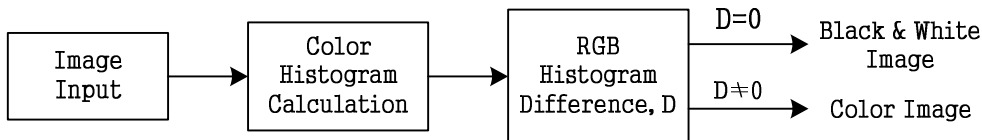


그림 1. 데이터베이스 생성 과정



where, $D = \sum |h(r) - h(g)| = \sum |h(g) - h(b)| = \sum |h(r) - h(b)|$

그림 2. 컬러와 흑백 영상 분류

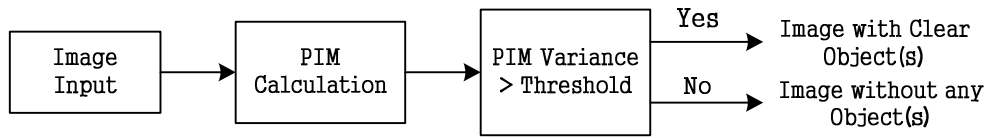


그림 3. 뚜렷한 개체를 가지는 영상과 그렇지 않은 영상 분류

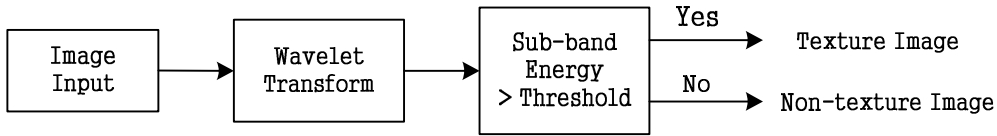


그림 4. 질감 영상과 비 질감 않은 영상 구분

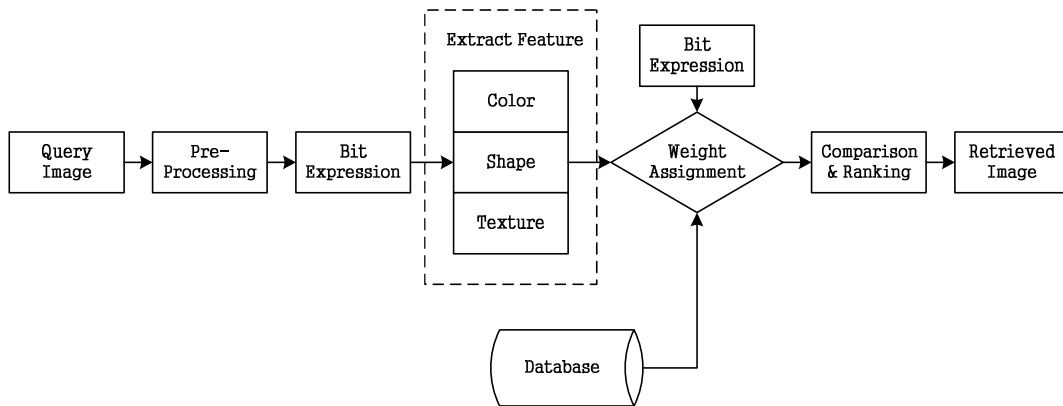


그림 5. 키 프레임 검색 과정

구체적인 가중치 부여 방법은 질의 영상의 컬러, 형태, 질감에 대한 3 비트와 데이터베이스 내 키 프레임의 3 비트를 비교해서 서로 같은 성질의 영상 특징을 가지면 가중치를 높게 부여하고 다를 경우에는 상대적으로 낮게 부여한다. 뚜렷한 개체를 가지는 경우(S), 컬러인 경우(C), 질감 영상인 경우(T)에는 각각 1이라는 비트를 할당하고 그렇지 않은 경우는 0이라는 비트를 할당한다. 질의 영상에 대한 형태, 컬러, 질감에 대한 분류 특성을 나타내는 비트는 각각 S_q , C_q , T_q 이고 데이터베이스 프레임에 해당되는 3 비트를 각각 S_d , C_d , T_d 라 하면 가중치는 0보다 큰 3의 배수 N 을 사용하여 각 특징에 대한 배타적 논리곱의 결과에 따라 상대적 가중치를 부여한다. 이 과정을 식 (8)의 형태로 나타낼 수 있다.

Step 1: $Query\ Image \rightarrow S_q : C_q : T_q$, (8)

$Data\ Image \rightarrow S_d : C_d : T_d$

Step 2: $S_q \odot S_d : C_q \odot C_d : T_q \odot T_d$
 $= a : b : c \rightarrow a : \beta : \gamma (a, b, c = 0\ or\ 1)$

Step 3: $If\ K \% 3 = 0 \rightarrow K, K > 0$

만약 모두 동일한 비트를 가지는 경우는 3가지 영상 특징 모두에 동일한 가중치 N 을 부여하게 되고 서로 다를 경우 1의 개수 또는 0의 개수에 따라서 $K \pm x$ 의 값을 갖게 한다. 이 과정을 식 (9)의 형태로 나타낼 수 있다. 그래서 형태, 컬러, 질감에 대한 가중치를 α, β, γ 로 나타낼 수 있고 영상에 따라 이 값들이 적응적으로 바뀌게 된다.

$$S_q \odot S_d : C_q \odot C_d : T_q \odot T_d \tag{9}$$

$$= \begin{cases} 1:1:1 \rightarrow K : K : K \\ 1:1:0 \rightarrow (K + \frac{K}{3}) : (K + \frac{K}{3}) : (N - \frac{2 \cdot K}{3}) \\ 1:0:1 \rightarrow (K + \frac{K}{3}) : (N - \frac{2 \cdot K}{3}) : (K + \frac{K}{3}) \\ 1:0:0 \rightarrow (K + \frac{2 \cdot K}{3}) : (K - \frac{K}{3}) : (K - \frac{K}{3}) \\ 0:1:1 \rightarrow (K - \frac{2 \cdot K}{3}) : (K + \frac{K}{3}) : (K + \frac{K}{3}) \\ 0:1:0 \rightarrow (K - \frac{K}{3}) : (K + \frac{2 \cdot K}{3}) : (K - \frac{K}{3}) \\ 0:0:1 \rightarrow (K - \frac{K}{3}) : (K - \frac{K}{3}) : (K + \frac{2 \cdot K}{3}) \\ 0:0:0 \rightarrow K : K : K \end{cases}$$

위와 같은 과정으로 각 특징들에 부과된 가중치 α , β , γ 가 자동적으로 결정되었다면 전체 유사도 T_SIM 은 식 (10)의 형태로 표현된다.

$$T_SIM = \alpha \times S_c(I, Q) + \beta \times S_s(I, Q) + \gamma \times S_t(I, Q) \quad (10)$$

기존 방법들은 α , β , γ 를 모두 1로 설정하거나 실험적 결과로서 성능이 가장 우수한 가중치를 고정시켜 사용한다. 또는 특정 검색 시스템에서는 사용자가 이 가중치를 직접 입력하도록 되어있는 경우도 있지만 최적의 검색 결과를 얻기 위해 이 값들을 적합하게 설정하기가 어렵다.

5. 실험결과

본 논문에서 제안한 적응적 가중치 부여 방법의 성능을 기존 고정 가중치 부여 방법과 비교하였으며 성능 평가 방법으로는 현재 널리 사용되고 있는 Precision vs. Recall, Retrieval Efficiency, Ranking Measure 등을 사용하였다[8].

실험 환경
 데이터베이스 내의 키 프레임 개수 : 3,200개
 (비디오 파일 개수 : 95개)
 윈도우 크기(T) : 15
 Precision vs. Recall 평가에 사용된 질의 영상의 개수 : 25개
 Ranking Measure 평가에 사용된 질의 영상의 개수 : 65개

위 실험에 사용된 질의 영상이 컬러지만 전체적으로 검은색과 흰색의 분포가 많다. 그래서 고정 가중치 방법을 사용한 경우 응답 프레임들은 형태적으로는 질의 영상과 아주 상이하지만 컬러 분포가 비슷하

여 전체 유사성이 높게 나타났다. 그러나 적응적 가중치 방법을 이용하면 고정 가중치 방법에서 발생하는 이런 문제를 보완할 수 있다. 표 1의 결과를 근거로 해서 가중치를 적응적으로 결정하면 효율적인 검색을 할 수 있다. 표 1에서 컬러, 형태, 질감 각각에 대해 결정된 가중치는 C_w , S_w , T_w 이고, C_SIM , S_SIM , T_SIM 은 각 특징에 대한 유사도를, C_pre , S_pre , T_pre 는 전처리 결과를 나타낸다.

질의 영상은 컬러 영상이고 원하지 않는 영상들은 흑백 영상이므로 컬러에 대한 가중치가 줄어들고 상대적으로 형태적 특징에 가중치가 높아지면서 유사성이 적은 키 프레임들이 윈도우 밖으로 밀려나게 된다

5.1 Retrieval Efficiency 결과

앞 절에서와 동일한 방법으로 25개의 서로 다른 질의 영상에 대해서 실험한 후 Retrieval Efficiency를 표 2에 나타내었다. 표 2에서 Data Type은 질의 영상의 순서를 나타내고 Query Data Number는 데이터베이스 내에서 추출한 질의 영상의 실제 번호를 나타낸다. The Number of Target Data는 질의 영상과 유사한 키 프레임들의 개수를 나타내고 The Number of Correct Retrieval Data는 주어진 윈도우 내에서 올바르게 검색된 영상 프레임의 개수를 나타낸다. 총 25개의 질의 영상 중 고정 가중치 방법에 비해 제안한 적응적 가중치 방법은 10개 데이터에서 검색 효율의 상승이 있었고 4개 데이터에서는 검색 효율이 떨어졌으며 나머지 데이터에서는 모두 동일한 성능을 나타내었다. 따라서 평균적으로 고정 가중치 방법은 0.65, 적응적 가중치 방법은 0.71의 검색

표 1. 검색 결과에 의해 할당된 가중치

Rank	Data	C_pre	S_pre	T_pre	C_SIM	S_SIM	T_SIM	C_w	S_w	T_w
1	197	1	218	2255	0	0	0	6	6	6
2	199	1	218	2254	0.000294	0.000169	0.000056	6	6	6
3	137	1	268	10192	0.130942	0.052337	0.001167	6	6	6
4	198	1	252	2647	0.219022	0.027100	0.002320	6	6	6
5	256	1	283	4384	0.141342	0.133131	0.014219	6	6	6
...
12	146	1	332	9014	0.243822	0.141091	0.030969	6	6	6
13	156	1	242	6782	0.214783	0.200373	0.003576	6	6	6
14	254	1	300	4429	0.266208	0.152947	0.011765	6	6	6
15	196	1	83	1320	0.285896	0.144648	0.001695	6	6	6

표 2. 25개의 질의 영상에 대한 검색 효율

data type	query data number	no. of target data	fixed weighting		adaptive weighting	
			no. of correct retrieval data	retrieval efficiency	no. of correct retrieval data	retrieval efficiency
1	10	24	9	0.45	9	0.45
2	16	11	6	0.55	9	0.82
3	32	7	5	0.71	5	0.71
4	94	26	20	1.0	20	1.0
5	107	23	16	0.8	16	0.8
...
20	749	6	2	0.33	2	0.33
21	822	10	2	0.2	2	0.2
22	885	14	10	0.71	10	0.71
23	899	11	6	0.55	5	0.45
24	936	13	7	0.54	7	0.54
25	958	9	4	0.44	4	0.44
-	Avg.	-	-	0.65	-	0.71

효율을 보여 줌으로써 제안한 방법이 더 좋은 성능을 나타냄을 알 수 있다.

5.2 Precision vs. Recall 결과

Precision vs. Recall의 관계 그래프는 Retrieval Efficiency와 유사한 측면이 있지만 성능의 변화를 좀 더 자세히 확인할 수 있다. 특히 그림 6과 같이 검색효율은 동일하지만 Precision vs. Recall의 관계 그래프가 서로 다를 수 있다는 것과 데이터 프레임의 순위 변화가 어떻게 바뀌어 가는지를 알 수 있다. 그림 6은 Data Type 3의 경우 예를 보인 것이다. 이 프레임에 대해서 검색 효율의 결과는 표 2에서 동일한 검색 효율을 보인다. 하지만 그림 6에서 나타나는 것처럼 그 성능이 고정 가중치 방법보다 제안한 적응적 가중치 방법에서 좀 더 이상적인 Precision vs. Recall의 그래프에 가까워짐으로써 성능이 향상됨을 보인다.

그림 7은 질의 영상 25개에 대한 평균적인 Precision vs. Recall의 결과를 나타내며 제안한 방법의 성능이 우수함을 볼 수 있다.

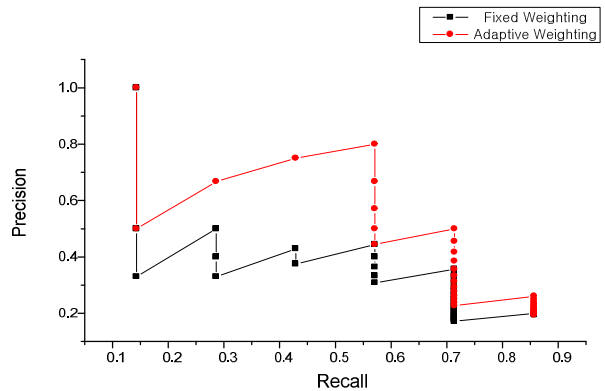


그림 6. Precision vs. Recall 그래프 (Data Type 3)

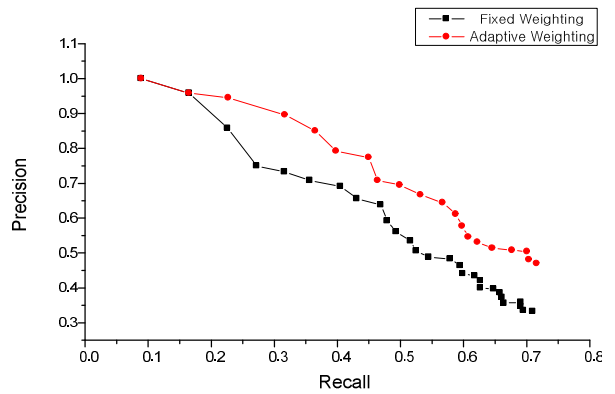


그림 7. 평균 Precision vs. Recall

5.3 Ranking Measure 결과

질의 영상에 대해 오직 한 개의 가장 유사한 키 프레임의 순위를 알아보기 위해 데이터 쌍을 구성된 후 각 영상들의 쌍에 대해 고정 가중치 방법과 적응적 가중치 방법을 적용한 결과는 표 3과 같다.

표 3에서 적응적 가중치 방법이 고정 가중치 방법에 비해 순위의 상승을 보였다. 하지만 질의 영상에 대해 유사 키 프레임이 카메라 줌 인, 줌 아웃 등의

표 3. 질의 영상에 대한 유사 프레임의 순위

	(a)	(b)	(c)	(d)
Fixed Weighting	43	59	49	31
Adaptive Weighting	18	12	29	32

표 4. Ranking Measures에 대한 성능(Q=65)

	Fixed Weighting	Adaptive Weighting
avg. r-measure	15.167	10.832
avg. p ₁ -measure	0.799	0.872

글로벌한 카메라 움직임이 포함된 경우에는 성능이 약간 열세이지만 평균적으로 봤을 때 average r-measure와 average p₁-measure값은 1에 가까울수록 우수한 성능임을 나타낸다. 질의 영상 65개에 대한 average r-measure, p₁-measure의 결과는 표 4와 같다.

위 결과를 살펴보면, 제안한 적응적 가중치 방법이 고정 가중치 방법에 비해 avg. r-measure와 avg. p₁-measure에서 좀 더 1에 가까운 10.822, 0.872의 값을 보임으로써 성능이 향상됨을 알 수 있다.

6. 결 론

본 논문에서는 키 프레임 내 여러 특징들이 결합해서 사용될 때 각 특징에 대한 가중치를 적응적으로 부여해서 사용자의 부담을 줄이고 검색 성능을 개선시킬 수 있는 새로운 방법을 제안하였다. 제안한 방법을 3,200개 키 프레임들로 구성된 데이터베이스에서 서로 다른 25가지 질의 영상 및 65개의 유사 영상의 쌍에 대해 실험하였다. 그리고 Precision vs. Recall, Retrieval Efficiency, Ranking Measure 등의 성능 평가 방법을 통해 제안한 고정가중치 부여 방법이 적응적 자동 가중치 부여 방법과 비교하여 검색 성능이 개선됨을 볼 수 있었다. 본 논문에서 얻은 연구결과는 비디오 데이터베이스의 관리 및 액세스, 비디오 클립에서 유사 장면 분류 등 대용량 비디오 검색 시스템에 다양하게 응용될 수 있으리라 기대된다.

참 고 문 헌

[1] L.Y. Duan, M. Xu, Q. Tian, S.S. Xu, and Jin, J.S, "A Unified Framework for Semantic Shot Classification in Sports Video," *IEEE Trans. on Multimedia*, Vol. 7, No. 6, pp. 1066-1083, 2005.
 [2] Guozhu Liu and Junming Zhao, "Key Frame Extraction from MPEG Video Stream," *Proc. Int. Symposium on Information Processing*, pp. 423-427, 2010.

[3] P.S. Hiremath and Jagadeesh Pujari, "Content Based Image Retrieval Using Color, Texture and Shape Features," *Proc. Int. Conf on Advanced Computing and Communications*, pp.780-784, 2007.
 [4] E. Saber and A.M. Tekalp, "Integration of Color, Edge, Shape, and Texture Features for Automatic Region-based Image Annotation and Retrieval," *Journal of Electronic Imaging*, Vol. 7, No. 3, pp. 684-700, 1998.
 [5] Zhu-Lin Tao, Xianglin Huang, and Xi Wang, "Shot Boundary Detection Based on Macro-block and DC Image," *Proc. Int. Conf on Management and Service Science*, pp. 1-3, 2009.
 [6] 김태희, 이규동, 정동석, "영상 데이터 베이스에서 PIM을 이용한 엔트로피 기반 색인 기법에 관한 연구," 한국통신학회 하계 학술발표대회 논문집, pp. 542-545, 1997.
 [7] Alexander Dimai, "Assessment of Effectiveness of Content Based Image Retrieval Systems," *Proc. Int. Conf. on Visual Information and Information Systems*, pp. 525-532, 1999.
 [8] Kang-Wook Kim and Seong-Geun Kwon, "New Framework for Automated Extraction of Key Frames from Compressed Video," *Journal of Korea Multimedia Society*, Vol. 15, No. 6, pp. 693-700, 2012.



김 강 욱

1992년 3월~1996년 2월 경북대학교 전자공학과 학사
 1996년 3월~1998년 2월 경북대학교 전자공학과 석사
 1998년 3월~2002년 2월 경북대학교 전자공학과 박사

2002년 3월~현재 삼성전자 무선통신연구소 수석연구원
 관심분야: 영상통신, 영상신호처리, 임베디드 시스템