

비디오 검색을 위한 시간 색상 상관관계그래프

박호식*, 이영식*, 김진한**, 나상동***, 배철수*,

*관동대학교 전자통신공학과

** (주) KT

***조선대학교 컴퓨터공학부

e-mail : mediana@netsgo.com

Temporal Color Correlograms for Video Retrieval

Ho-Sik Park*, Young-Sik Lee*, Jin-Han Kim**,

Sang-Dong Na***, Cheol-Soo Bae*

*Dept of Electronics Communications, Kwandong University

**Korea Telecom Corporation

***Dept of Computer Engineering, Chosun University

요 약

본 논문은 분할된 비디오 화면들의 색상 내용을 기반으로 하는 새로운 영상 검색 방법을 제안 하고자 한다. 제안된 시간 색상 상관그래프는 공통적인 통계 데이터를 이용하여 비디오 화면 내의 공간-시간 관계를 계산한다. 시간 색상 상관 그래프는 내용 기반의 영상 검색에 매우 효과적인 것으로 밝혀진 HSV(Hue, Saturation, Value) 색상 상관 그래프를 기반으로 하고 있다. 시간 색상 상관 그래프는 하나의 비디오 화면으로부터 추출된 프레임 샘플의 양자화 된 HSV 색상 값의 자기상관관계를 이용하여 구성하였다. 본 논문에서는 11시간 분량의 분할된 MPEG-1 비디오에 대한 질의와 질의에 대한 관련성 판정을 하고자 내용 기반의 멀티미디어 검색 시스템을 구축하여 실험하였다. 실험 결과 제안된 방법이 시각 정보만을 필요로 하는 검색에 있어 기존의 다른 검색 방법보다 우수한 결과를 나타냄을 증명하였다.

1. 서론

내용 기반 영상 검색은 1990년대 초반부터 활발하게 연구[3]가 진행되어 오고 있으며, 많은 실험적인 영상 검색 시스템들이 연구되고 있다[5,6,8,9,11]. 색상을 기반으로 한 내용 검색은 처음에는 평균 색상과 같은 간단한 통계로부터 점차 색상 히스토그램과 같은 공간적인 색상 특성[9,5,8,11,12,13]으로 발전하였으나 색상 히스토그램은 색상의 공간적 구조와 관련하여 제한적인 분간 능력을 갖고 있을 뿐이었다. 그러나, 색상 상관그래프와 자기상관그래프는 고정된 픽셀 거리에서 색상 조합을 발견할 수 있는 확률을 계산하여 효율적인 공간 검색 기능을 제공한다. Huang[12]은 히스토그램과 색상 일관성 벡터와 비

교하여 자기상관그래프가 매우 뛰어난 검색 성능을 제공함을 입증하고, 검색 방법[14,15]을 개발하였다.

본 논문에서는 HSV 색상 공간의 여러 다른 양자화방법들을 검토하여, 상관그래프의 색상 내용의 변화에 대한 민감도와 다양한 조명 조건에 대한 탄력성을 개선하고자 하였다. 본 논문은 검색을 위한 시간적 비디오 특성을 생성하기 위해 HSV 컬러 상관 그래프를 확장하였고, 제안된 시간 색상 상관그래프(TCC: Temporal Color Correlogram)는 Ojala가 제시한 HSV 색상 히스토그램과 HSV 색상 상관 그래프를 이용하는 화면의 정적 핵심프레임 표현과 대비하여 평가하였다. 먼저 비디오 세그먼트들을 저장하고 특성들을 계산한 다음 시스템에 샘플 세그먼트들을 제공함으로써 그 세그먼트들을 검색하였다.

2. 시간 색상 상관 그래프

2.1 비디오 구조

디지털 비디오의 근본적인 요소는 프레임(I 로 표시)이다. 디지털 비디오 분석에서, 녹화 장비의 기계적 한계 내에서 연속적이고 일관적인 움직임을 표현하는 최장의 일련의 프레임들을 하나의 화면(S 로 표시)이라고 한다. 일관적인 동작의 흐름과 같은 화면의 특성들은 그 화면에 대한 '기계적인' 정보를 담고 있으며, 비디오 색인을 위한 분석 알고리즘에서 활용된다.

2.2 히스토그램, 상관그래프 및 자기상관그래프

단일 프레임으로부터 상관그래프를 계산하기 위한 수학적 원리는 다음과 같다. 픽셀 $p(x,y)$ 로 구성된 $X \times Y$ 이미지를 I 라고 하고, $[C]$ 를 그 이미지 내에서 일어날 수 있는 색상 c_1, \dots, c_c 를 원소를 하는 집합 C , $[D]$ 를 거리 d_1, \dots, d_D 를 원소를 하는 집합 D 라고 한다. 여기서 거리들은 L_∞ 을 이용하여 측정되어진다. 픽셀 p 에 대해서, $I(p)$ 는 픽셀 p 의 색상 c 를 나타내며, I_c 가 픽셀 p 에 대응한다면, p 에 대해 $I(p)=c$ 라고 할 수 있다. 그리고 I 의 히스토그램과 상관그래프는 색상 c_i 와 색상 조합 (c_i, c_j) 에 대해 각각 다음과 같이 정의된다.

$$h_{c_i}(I) \equiv \Pr [p \in I_{c_i}] \quad (1)$$

$$\gamma_{c_i, c_j}^{(d)}(I) \equiv \Pr_{p_1 \in I_{c_i}, p_2 \in I} [p_2 \in I_{c_j} \mid |p_1 - p_2| = d] \quad (2)$$

여기서 수식 (1)은 I 에 있는 어떤 픽셀이 색상 c_i 가 될 확률이며, 수식 (2)는 주어진 어떤 픽셀 p_1 의 색상이 c_i 일 때 주어진 픽셀 p_1 로부터 d 거리에 있는 픽셀 p_2 가 색상 c_j 일 확률이다. 이 색상들 사이의 공간적 관계는 히스토그램에서는 나타낼 수 없지만 상관그래프에서는 나타낼 수 있다. 상관도표의 크기는 $C^2 D$ 이다. 이미지 I 의 자기상관그래프는 거리 d 에서 동일한 색상들(수식 2에서 $c_j=c_i$)를 발견할 확률이며, 이것은 상관그래프에 비교하여 연산의 이점이 있다. 본 논문에서, HSV 컬러 상관도표는 수식 2를 이용하여 $c_j=c_i$ 로 계산하였으며, HSV 색상 히스토그램 특성은 수식 1을 이용하여 계산하였다.

HSV 컬러 히스토그램에 대해서, 가중 유클리드 거리법을 이용하여 질의 핵심프레임 Q 와 참조 핵심프레임 R 사이의 차이를 측정했으며, 자기상관그래프의 경우, Huang 등의 연구[12]에 따라 L_1 norm을

이용하여 비유사성을 측정하였다. 그 공식들은

$$D_h(Q, R) = [h(Q) - h(R)]' A [h(Q) - h(R)] \quad (3)$$

$$D_a(Q, R) = \sum_{c \in [C], d \in [D]} |\alpha_c^{(d)}(Q) - \alpha_c^{(d)}(R)| \quad (4)$$

수식(3)에서 A 는 가중치 행렬로서, 그 원소들 a_{ij} 는 색상 c_i 와 c_j 사이의 유사성이며, $h(Q)$ 는 양자화된 색상 집합 $[C]$ 를 가진 영상 Q 의 히스토그램을 산출해낸다. HSV를 고려할 때, 컬러들의 유사성은 컬러 공간에서의 3차원 색상 포인트들의 역거리로서 계산된다. 수식(4)에서 Q 는 질의 핵심프레임, R 은 참조 핵심프레임이며, $\alpha_c^{(d)}(I)$ 는 영상 I 내에서 색상 c 에 대한 자기상관그래프 값을 거리 d 를 이용하여 나타낸 것이다 [1].

2.3 시간 영역으로의 확장

비디오 화면 내용의 시간적 변화는 시간 상관그래프를 이용하여 설명할 수 있다. 색상 히스토그램과 같은 종래의 접근방법과 비교할 때 이점은 사소한 공간적 색상 관계들의 시간적 변화를 포착할 수 있다는 것이다. 그림 1은 시간적 색상을 보여준다. 색상 히스토그램이 이들 프레임 내의 색상의 비례적 양을 나타내는데 반해 시간 상관그래프는 시간이 지나면서 생기는 이 색상들의 공간적 변화들에 대한 정보를 포착한다.

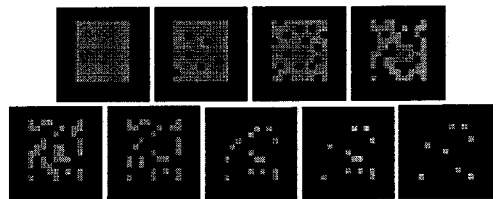


그림 1. 프레임 순서에 의해 나타나는 시간적 색상 변화. 시간 상관그래프는 색상 요소들의 확산을 포착하는데 반해, 히스토그램을 그릴지 못하다.

N 을 화면 S 로부터 추출된 샘플 프레임 I^n 의 수라고 하고, 여기서 n 의 값은 샘플 프레임 순서 색인에 따라 1에서 N 까지 될 수 있다. 시간 상관그래프는 다음과 같이 계산되어진다.

$$\bar{\gamma}_{c_i, c_j}^{(d)}(S) \equiv \Pr_{p_1 \in I_{c_i}^n, p_2 \in I^n} [p_2 \in I_{c_j}^n \mid |p_1 - p_2| = d] \quad (5)$$

이 수식은 색상 c_i 의 픽셀 p_i 가 주어졌을 때, 그 주어진 픽셀 p_0 로부터 d 거리에 그 화면의 샘플 프레임 I^n 중 색상 c_j 가 있을 확률이다.

본 논문에서 사용된 시간 색상 상관 그래프는 c_j 를 c_i 로 대치함으로써 수식2에서 얻은 바와 같은 자기상관 그래프로 계산된다.

3. 실험 결과 및 고찰

본 논문에서는 11 시간 분량의 MPEG1 비디오를 사용하여 실험을 진행하였다. 11시간 분량의 비디오를 이용하여 7,375개의 화면을 분할 생성하였고, 생성된 분할 화면들의 평균 길이는 5초였다. 분할한 화면프레임들로부터 첫번째 프레임이 대표 핵심프레임으로 선택되었으며, 정적 이미지 특성들인 색상 상관 그래프와 HSV 색상 히스토그램이 계산되었다. 시간 색상 상관 그래프를 계산하고 다양한 화면 길이에 대해 샘플들의 수를 동일하게 유지하기 위해서, 각 화면은 고정된 샘플링 지점으로 샘플링된 프레임의 수가 40을 초과하지 않도록 하였다. 최종적으로 화면 특성들이 내용 기반의 멀티미디어 검색 시스템(CMRS: Content-based Multimedia Retrieval System)을 구성하여 실험하였다.

검색 사례를 기초로 하여 총 74개의 주제를 정하였고, 각 주제는 다양한 일반적 항목들을 포함하고 있다. 데이터베이스 내에서 주제에 병합하는 항목을 식별하지 못하는 경우를 일반적인 검색으로 간주하였고, 검색 시스템에서 사용되는 검색 단서들은 흔히 텍스트 설명을 추가한 비디오, 오디오 그리고/또는 영상 샘플들을 갖고 있지만, 제안된 시스템에서는 단지 시각적 정보만을 사용하였다. 주제들은 질의 절차에서의 검색자의 참여 정도에 따라 자동적, 상호적, 또는 자동적+상호적으로 분류되었다. 검색 결과 평가는 실험자가 육안으로 클립이 그 주제와 관련성이 있는지 여부를 판단하였다.

실험에 사용한 비디오의 첫번째 구성은 포함한 주제 검색을 위해 시간 색상 상관 그래프를 사용해서 얻었으며, 둘째 구성은 색상 상관 그래프를 사용하였다. 서로 다른 주제 범주에서 시간 색상 상관 그래프 특성들을 이용한 검색으로부터 얻은 최상의 100개의 결과들을 사용한 평균 정확도는 0.08(상호적 주제에 대해), 0.13(자동적 및 상호적 주제에 대해) 그리고 0.24(자동적 주제에 대해)였다. 이 결과는 순전히 자동적 방법으로서 시간 색상 상관 그래프가

자동적+상호적으로 분류된 주제에서는 성능이 떨어짐을 보여준다. 그 이유는 색상 구조 특성 외에 어떤 다른 단서들도 질의에 사용되지 않았기 때문이다. 일반 검색 결과들은 뛰어난 성능을 보이지는 않았지만, 자동적 주제에서 평균 정확도는 높게 나타났다.

기존 항목 검색 결과는 검색된 화면이 기존 항목과 성공적으로 일치하는 것으로 정의되는 4개의 서로 다른 매개변수들을 이용하여 측정하였다. 일치에 있어 낮은 기준은 그 검색된 화면의 시간 길이의 최소한 3분의 1이 기존 항목 세그먼트와 동일한 규칙을 갖는 것으로 하였으며, 높은 기준은 화면 길이의 3분의 2가 겹치는 것으로 하였다.

본 논문에서 겹침에 대한 최소 요구 조건 없이 화면들이 겹치는 한 일치하는 것으로 인정하여 알려진 항목아이템 결과를 제시하였다. 기존 아이템과 질의 세그먼트가 겹칠 때 사용자가 그 비디오의 정확한 지점을 파악하기 위해 검색된 세그먼트들을 손쉽게 재생할 수 있었다. 이는 화면의 평균 길이가 5초이기 때문이다. 겹침 기준을 적용하지 않은 가장 성공적인 주제에 대한 100개의 최상의 결과들 내에서 정확도 곡선은 그림2에 나타내었다.

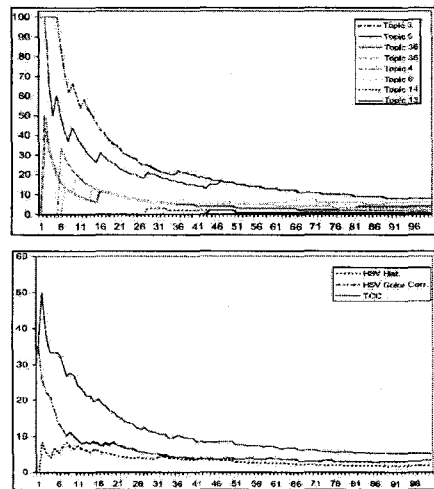


그림2. 시간 색상 상관 그래프 특성(상단)을 이용한 최상의 정확도 곡선들과 서로 다른 특성들을 이용한 평균 정확도 곡선들 (하단)

그림2는 또한 모든 테스트된 특성들, 즉 HSV 히스토그램, HSV 색상 상관 그래프 및 시간 색상 상관 그래프에 대한 평균 정확도 곡선들을 보여주고 있다. 이 그림에 따르면 시간 색상 상관 그래프는 정

적 핵심프레임에 기초한 접근 방법들에 비해 성능이 월등하였다.

결과들로부터 HSV 색상 히스토그램 결과가 정확하지 않을 때에도 시간 색상 상관 그래프는 조명 및 색상 채도에 영향을 받지 않으면서 유사하게 구성된 주제3에 대한 몇 가지 검색 결과들은 그림3에 나타내었다.

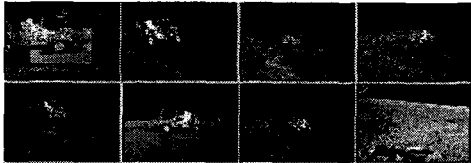


그림3. '달 표면을 운행하는 달 표면 자동차를 찾아라' 질의에 대한 상위 8개의 결과들. 가장 유사한 화면은 상단 맨 왼쪽에 있는 것이며, 유사도는 위에서 아래로 왼쪽에서 오른쪽으로 갈수록 줄어든다.

4. 결론

본 논문에서 제안된 방법이 시각 정보만을 필요로 하는 작업에서는 높은 결과를 보여주고 있지만 일반적인 주제 검색에서 시간 색상 상관 그래프와 색상 상관 그래프에 대한 차이가 많이 나타났다. 이것은 오디오 및 텍스트를 검색 단서에 포함시킴으로써 개선할 수 있다고 생각된다. 제안된 시스템은 각 주제에 대해 최대 100개의 최상의 세그먼트들을 찾아냈는데, 이것은 기존의 영상 검색 시스템들과 비교할 때 매우 높은 정확도를 가짐을 증명하였다.

참고문헌

[1] Ojala, T, Rautiainen, M, Matinmikko, I & Aittola, M, (2001), "Semantic image retrieval with HSV correlograms." Proc. 12th Scandinavian Conference on Image Analysis, Bergen, Norway: pp.621-627.

[2] TREC-2001 Video Retrieval Track Home Page, (10/25/2001). "http://www-nlpir.nist.gov/projects/t01v/t01v.html"

[3] Eakins, J & Graham, M, (1999), "Content-Based Image Retrieval: A report to the JISC Technology Applications Programme." Institute for Image Data Research, University of Northumbria at Newcastle, United Kingdom. http://www.unn.ac.uk/iidr/research/cbir/report.html.

[4] Ojala, T, Kauniskangas, H, Keränen, H, Matinmikko, E, Aittola, M, Hagelberg, K, Rautiainen, M & Häkkinen, M (2001), "CMRS : Architecture for content-based multimedia retrieval." Proc. Infotech Oulu International Workshop on Information Retrieval, Oulu, Finland: pp.179-190.

[5] Ogle, V & Stonebraker, M, (1995), "Chabot: retrieval from a relational database of images." IEEE Computer Magazine 28: pp.40-48.

[6] Ortega, M, Rui, Y, Chakrabarti, K, Porkaew, K, Mehrotra, S, Huang, TS, (1998), "Supporting ranked Boolean similarity queries in MARS." IEEE Transactions on Knowledge and Data Engineering 10: pp.905-925.

[7] Darwish, K, Doermann, D, Jones, R, Oard, D, and Rautiainen, M, (2001), "Experiments at Maryland: CLIR and Video." Text REtrieval Conference TREC 10.

[8] Pentland, A, Picard, R & Sclaroff, S, (1996), "Photobook: content-based manipulation of image databases." International Journal of Computer Vision 18: pp.233-254.

[9] Flickner, M, Sawhney, H, Niblack, W, Ashley, J, Huang, Q, Dom, B, Gorkani, M, Hafner, J, Lee, D, Petkovic, D, Steele, D & Yanker, P, (1995), "Query by image and video content: The QBIC system." IEEE Computer Magazine 28: pp.23-32.

[10] Foley, J, van Dam, A, Feiner, S & Hughes, J, (1990), "Computer Graphics: Principles and Practice." Addison Wesley Publishing, 1200.

[11] Smith, J & Chang, S-F, (1996), "VisualSEEK: a fully automated content-based image query system." Proc. Fourth ACM International Conf. on Multimedia, Boston, MA: pp. 87-98.

[12] Huang, J, Kumar, SR, Mitra, M & Zhu, WJ, (1997), "Image indexing using color correlograms." Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico: pp.762-768.

[13] Pass, G, Zabih, R & Miller, J, (1996), "Comparing images using color coherence vectors." Proc. Fourth ACM International Conference on Multimedia, Boston, MA: pp.65-73.

[14] Huang, J, Kumar, SR, Mitra, M & Zhu, WJ, (1998), "Spatial color indexing and applications." Proc. Sixth International conference on Computer Vision, Bombay, India: pp.602-607.

[15] Huang, J, Kumar, SR & Mitra, M, (1997), "Combining supervised learning with color correlograms for content-based image retrieval." Proc. Fifth ACM International Conference on Multimedia, Seattle, WA: pp.325-334.