

색상 히스토그램과 특징점 추출 알고리즘을 활용한 장면 전환 검출 방법

오현주^{1*}, 고완진¹, 박지용¹¹구미전자정보기술원 실감미디어연구센터

*hjoh@geri.re.kr

Scene Change Detection Method using Color Histogram and Feature Detection Algorithm

Hyunju Oh¹, Wanjin Ko¹, Jiyong Park¹¹Realistic Media Research Center,

Gumi Electronics & Information Technology Research Institute

요 약

장면 전환 검출에서 단일 특성을 사용하는 경우 발생 가능한 정확도 감소의 문제를 해결하기 위해 색상 히스토그램 분포 차 분석과 특징점 추출 알고리즘을 활용한 방법을 제안한다.

1. 서론

통신 및 컴퓨팅 기술의 발전에 따라 대용량의 문자, 음성, 영상 등의 많은 멀티미디어 정보가 생산되고 있다. 대용량 멀티미디어를 보다 효율적으로 검색하고 관리하는 것이 중요한 문제로 대두되었고[1], 특히 의미 있는 장면들로 분류 및 검색하는 장면 전환 검출 기술은 영상 정보를 효과적으로 관리하고 검색하는데 중요한 분야 중 하나로 인식되고 있다.

2. 장면 전환 검출 방법

장면 전환 검출은 영상에서 연속적으로 주사되는 각 프레임에 담겨있는 정보를 이용해 장면이 전환되는 지점을 찾아내는데 이용되는 기술이다. 장면 전환은 급진적 장면 전환과 점진적 장면 전환으로 구분할 수 있다. 급진적 장면 전환은 화면 내 전체적인 변화가 생긴 경우이며, 점진적 장면 전환은 유사하게 변화하는 경우에 해당한다. 급진적 방법은 쉽게 검출이 가능하지만 점진적 방법은 임계치 설정값에 따라 높은 경우 너무 많은 장면 전환을 검출하게 되는 등의 한계가 있다.

3. 특징

영상에서 특징으로 분류될 수 있는 것은 색상, 질감, 모양 등이 있다. 색상의 경우 가장 널리 사용되는 특징으로 영상의 크기와 방향에 독립적인 특성을 지닌다. 일반적으로 사용하는 색상 특징으로는 컬러 히스토그램, SCD(Scalable Color Descriptor), CLD(Color Layout Descriptor), CSD(Color Structure Descriptor) 등이 있다. 질감 특징은 구름, 나무, 식물, 벽돌 등과 같은 표면

특성과 관련된 것으로 표면의 구조적인 배열과 주위 화소와의 관계에 대한 정보를 나타낸다. 질감 특징은 GLCM(Gray-level Co-occurrence Matrix)[2]를 이용한 통계적 질감 특징, EHD(Edge Histogram Descriptor), 웨이블릿 변환을 이용한 다 해상도 질감 특징 등이 있다. 모양 특징은 객체의 윤곽선을 나타내는 특징으로 인간이 시각적으로 이미지를 분류하거나 인식하는데 있어 가장 중요한 정보이다. 그러나 모양 특징은 배경이 복잡하거나 객체와 배경 구분이 모호한 경우 정확한 추출이 어렵다. 일반적으로 사용하는 모양 특징으로는 Fourier Descriptor[3], Chain Code[4], Zernike Moment[5] 등이 있다.

3.1. SIFT(Scale Invariant Feature Transform)

특징점 추출 기법은 SIFT와 SURF가 많이 사용되는데, SIFT 알고리즘은 크기, 방향, 조명 등에 불변한 특징을 찾아낸다[6]. SIFT의 전처리 과정 첫 번째 단계는 그림 1과 같이 Scale Space를 구축해 영상 자체를 다양한 크기로 변화시키면서 나타나는 크기에 불변한 특징을 추출하는 것이다. 두 번째 단계는 영상 보정을 위한 방법으로, 너무 밝거나 어두운 쪽으로 편중된 픽셀값 분포를 그레이 스케일 전체로 분포시킨다.



(그림 1) Scale Space 구축

특징점 추출을 위한 첫 번째 단계는 가우시안 영상을 생성하는 것이다. 가우시안 필터의 표준 편차 σ 를 다른 분산 값으로 사용하며 입력 영상의 흐림 정도를 (blurred) 증가시킨다.

DoG 영상은 다른 분산 값을 갖는 두 개의 가우시안 영상의 차를 말한다. 다른 분산 값을 이용해 가우시안 영상을 생성할 경우, 영상의 에지 부분에서 흐릿해지는 정도가 다르다. 따라서 가우시안 영상의 차를 이용하면 에지 기반의 영상이 생성되며 이를 이용하여 특징점 추출 과정이 이루어지며 DoG 영상과 계산은 각각 그림 2, 식 (3.1)과 같다.

$$D(x,y,\sigma) = G(x,y,k\sigma) - G(x,y,\sigma) \cdot I(x,y) \\ = L(x,y,k\sigma) - L(x,y,\sigma) \quad (3.1)$$



(그림 2) DoG 영상

다음 단계는 극값을 추출하고 에지 성분을 제거하는 단계이다. 현재 픽셀을 기준으로 총 26개의 주위 픽셀의 극값을 비교한다. 만약 현재 픽셀의 밝기 값이 26개 픽셀과 비교했을 때 최대 혹은 최소인 경우 특징점 후보로 결정된다. 이전 단계에서 특징점 후보들을 추출을 위해 DoG 영상을 사용하였다. 이 결과로 주로 에지 부근에서 특징점 후보들이 많이 추출됨과 동시에 주변 픽셀 값과 차이가 많이 나지 않는 점들 또한 추출된다. SIFT에서의 특징점은 코너 반응을 보이는 점들로 추출되므로 이러한 특징점 후보들은 제거되어야 한다. 주변 픽셀 값과 차이가 적은 점의 경우는 테일러 급수를 통하여 제거되고, 에지 부근의 경우는 헤이시안 행렬을 이용해 제거된다.

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \quad (3.2) \\ \hat{x} = \frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x}, D\hat{x} = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x}$$

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (3.3) \\ Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \\ Det(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \\ \frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r+1)^2}{r} \\ \frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r}$$

다음 단계는 Orientation Assignment이다. 객체의 매칭을 위해서는 회전에 강인한 특성을 지녀야한다.

따라서 특징점은 방향 정보를 가지고 있어야 한다. 방향 정보를 구하기 위해 특징점 주변 픽셀 값의 경사도를 계산하며 특징점의 경사도는 360도 범위에 10도 단위로 구분한 36개의 방향으로 표현한다.

3.2. SURF(Speeded Up Robust Features)

본 논문에서는 SIFT와 견줄만한 성능을 보임과 동시에 속도를 향상시킨 특징점 추출 방법인 SURF를 사용하였다. SURF의 경우 속도 향상을 위해 Integral Image를 사용하고 계산 과정이 간단한 근사화한 헤이시안 검출기를 사용한다. Integral Image는 네 개의 point만 알면 그 영역의 면적을 간단하게 바로 계산할 수 있다. 식 (3.4)는 x, y 위치에서 Integral Image $\Pi(x, y)$ 을 구하는 과정을 나타내며, $I(i, j)$ 는 원본 영상에서 i, j 위치의 픽셀 값을 의미한다.

$$\Pi(x,y) = \sum_{i=0}^x \sum_{j=0}^y I(i,j) \quad (3.4)$$

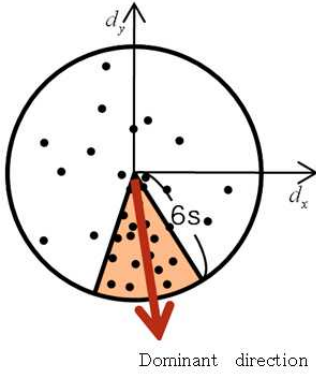
근사화된 헤이시안 검출기는 식 (3.5)과 같이 헤이시안 행렬식을 사용하며, 표현된 근사화한 필터를 이용한다. 헤이시안 검출기는 정수화 되지 않았기 때문에 연산량이 많고 시간 복잡도가 증가한다. x축, y축, xy축에 대해 네 개의 entry를 식 (3.5)를 이용해 계산하면 interest points 후보 지점을 검출할 수 있다.

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \quad (3.5)$$

$L_{xx}(x,\sigma)$ 는 x위치의 입력 영상과 σ 의 분산을 갖는 가우시안의 x방향 2차 미분 값 $\frac{\partial^2}{\partial x^2}g(\sigma)$ 과의 컨볼루션 값을 의미하고 $L_{xy}(x,\sigma)$, $L_{yy}(x,\sigma)$ 는 xy방향, y방향으로의 컨볼루션 값을 의미한다.

특징점 서술에서는 특징점을 중심으로 추출된 6s(Scale) 반경 내의 픽셀들에 대하여 가로 방향과 세로 방향으로의 하르 웨이블릿 특징을 계산한다. 그리고 슬라이딩 윈도우를 이용해 윈도우 내에 기울기들을 모두 더해 다수의 합 벡터를 생성한다. 특징점의 주방향은 그림 3의 붉은 선으로 표현된 것처럼 합 벡터가 가장 많이 가리키는 방향으로 할당된다.

특징점이 계산된 후 최종 단계로 서술자를 구성한다. 특징점을 중심으로 20×20 픽셀의 기울기를 계산하고, 계산된 기울기들은 dx, dy, $\sum dx$, $\sum dy$ 와 같이 4개의 특징 벡터를 만들어낸다. 이 때 dx와 dy는 각각 x와 y방향에 대한 기울기이다.



(그림 3) 주 방향 결정

4. SURF를 활용한 장면 전환 검출 방법

4.1. 색상 히스토그램 차 분석

첫 번째 단계는 색상 히스토그램 분석 단계로, 제안하는 방법은 현재 및 이전 프레임 간의 색상 히스토그램 분포 차를 이용한다. 식 (4.1)에 따라 유사도를 측정한다. $H_i(n)$ 과 $H_j(n)$ 은 각각 이전 프레임의 히스토그램과 현재 프레임의 히스토그램을 나타낸다. 이후 식 (4.2)에 해당하는 임계치 Difference threshold인 *Weighted average*와 비교한다. α 값은 실험 결과에 따라 0.6으로 사용하였다.

$$w(i, j) = \sum_{n=0}^{N-1} |H_i(n) - H_j(n)| \quad (4.1)$$

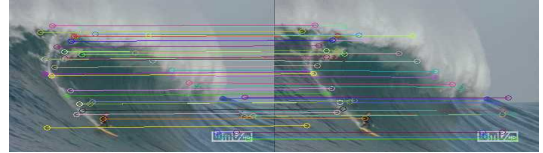
$$W_{avg} = W_{avg} \cdot \alpha + cur \cdot (1 - \alpha) \quad (4.2)$$

4.2. SURF를 활용한 장면 전환 검출

컬러 히스토그램 분포 차를 이용한 프레임 간 유사도 판별 후 장면 전환 후보들이 결정된다. 장면 전환 후보들에 한하여 SURF를 이용한 장면 전환 검출 단계가 진행되며 이 단계를 통해 정확한 검출을 하고자 하였다.

방법은 SURF는 feature 공간에서 가장 짧은 거리를 갖는 특징점 간 매칭을 수행하는데 이 거리는 식 (4.3)로 표현되는 유클리드 거리에 해당한다. 이 거리를 평균과 표준 편차 임계치와 비교하는데, 임계치 이상인 경우 장면 전환이 일어났다고 판단한다. 이 과정이 끝나면 다음 장면 전환 검출을 위해 각종 수치를 초기화하고 4.1절, 4.2절 과정을 반복한다.

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \\ = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (4.3)$$



(그림 4) SURF 특징점 매칭 과정

5. 실험 결과

성능 분석을 위해 색상 히스토그램 방법과 제안하는 방법 간의 실제 성능을 실험하였다. 개발 언어는 MATLAB, 실험은 CIF(352×288) 크기의 동영상 압축 표준 실험 영상 중 인물, 풍경 등 7개의 영상을 사용하였다.

<표 1> 실험에 사용한 영상 정보

분류	이름	재생 시간(초)
Documentary	Chionin	70
	Journey of Amazing Cave	150
Movie	Taxi 3	200
	Terminator 2	180
Nature	Amazon	240
	Coral	240
Sports	Step Into Liquid	120

5.1. 실험 결과 분석

기존 방법과 제안 방법의 비교를 위해 Taxi 3 영상의 일부분인 120 프레임을 비교해 보았다. 총 5번의 장면 전환이 일어나는 이 영상의 경우, 색상 히스토그램 단일 특성만 고려하는 경우 전환이 일어나지 않은 영상임에도 불구하고 장면 전환으로 검출되는 문제가 있었다. 제안한 방법은 정확하게 검출하는 것을 알 수 있었다.

5.1.1. 프레임 간 색상 히스토그램 분포 차 추출

CIF 크기의 압축 표준 영상들을 이용해 프레임 간 색상 히스토그램 분포 차를 계산하였고, 영상의 밝기, 색상 분포에 따라 히스토그램 분포 차가 달라지는 것을 확인할 수 있었다.

5.1.2. SURF 특징점 매칭 과정

두 번째 과정은 색상 히스토그램 분포 차를 이용해 장면 전환 후보들을 대상으로 수행되는 과정이다. 성능평가 척도로는 Recall, Precision, F-Measure를 사용하였고 정의는 식 (5.1)과 같다[7].

$$Recall = \frac{N_c}{N_c + N_m} \\ Precision = \frac{N_c}{N_c + N_f} \quad (5.1) \\ F-Measure = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

여기서 N_c , N_m 그리고 N_f 는 각각 검출된 장면 전환 수, 검출하지 못한 장면 전환 수, 잘못 검출된 장면 전환 수를 의미한다. 실험 데이터를 통해 Recall, Precision을 구하고 이를 이용해 조화 평균인 F-Measure를 구한다. 식 (5.1)에 따라 N_m 와 N_f 이 작을수록 좋은 성능을 나타낸다고 할 수 있다.

표 2는 색상 히스토그램과 제안한 방법의 Recall, Precision, F-Measure를 보여준다. 실험 결과 장면 전환이 한 번 일어나는 Chionin 영상에서는 두 방법 모두 정확하게 장면 전환을 검출하였다. 빠르게 변화하는 장면들로 이루어진 Terminator 2와 Amazon 영상에서 히스토그램만 고려한 경우 Recall과 Precision이 많이 차이나는 것을 확인할 수 있었다. 또한 점진적 장면 전환이 일어나는 Coral 영상의 경우 제안한 방법과 달리 너무 많은 장면 전환을 검출하는 것을 알 수 있었다.

<표 2> 장면 전환 검출 정확도 비교

분류	이름	Recall	Precision	F-Measure
Documentary	Chionin	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
	Journey	100.0 (100.0)	100.0 (86.6)	100.0 (92.8)
Movie	Taxi 3	90.2 (89.0)	81.6 (23.5)	85.7 (37.2)
	Terminator 2	87.1 (65.0)	88.3 (62.9)	87.7 (63.9)
Nature	Amazon	90.0 (80.5)	92.4 (84.2)	91.2 (82.3)
	Coral	84.3 (93.0)	75.0 (37.0)	79.4 (52.9)
Sports	Step	94.5 (71.8)	90.0 (66.1)	92.2 (68.8)

주 : () 안은 색상 히스토그램

사사

본 논문은 대한민국 정부(산업통상자원부 및 방위사업청) 재원으로 민군협력진흥원에서 수행하는 민군기술협력사업의 연구비 지원으로 작성되었습니다.(협약번호 UM2203RD2)

참고문헌

[1] Choi Hyun-Hwa, Lee Mi-Young, Lee Kyu-Chul, "A Scalable Index for Content-based Retrieval of Large Scale Multimedia Data," The Korea Contents Association, Vol. 7, pp. 726-730, May 2009.

[2] M. M. Mokji, S.A.R. Abu Bakar, "Adaptive Thresholding based on Co-Occurrence Matrix Edge Information." Journal of Computers, Vol. 2, No. 8, October 2007.

[3] D. Zhang, G. Lu, "A Comparative Study on Shape Retrieval using Fourier Descriptors with Different Shape Signatures," JVCIR, No. 14, pp. 41-61, 2003.

[4] Walid Shahab, Hazem Al-Otum, Farouq

Al-Ghoul, "A Modified 2-D Chain Code Algorithm for Object Segmentation and Contour Tracing," IAJIT, Vol. 6, No. 3, July 2009.

[5] Sun-Kyoo Hwang, Whoi-Yul Kim, "A Novel Approach To The Fast Computation of Zernike Moments," Pattern Recognition, Vol. 39, pp. 2065-2076, March 2006.

[6] David G. Lowe, "Distinctive Image Features from Scale Invariant Keypoints," International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, January 2004.

[7] Hyunju Oh, Jiyong Park, Seokmin Hong, Hyunmin Kang, "Scene Change Detection Based on SURF," ACK 2022, Vol. 29, No. 2, pp. 637-639, November 2022.