

영역 정보를 이용한 비디오 데이터의 카메라 모션 검출 및 대표 프레임 선택 방법

이용현*, 강행봉**, 박용진*
*한양대학교 전자공학과, **가톨릭대학교 컴퓨터 공학부

Camera Motion Detection and Key-Frame Selection from Region-Based Segmented Video Data

Yong-Hyun Lee*, Hang-Bong Kang**, Yong-Jin Park*
*Dept. of Electronic Engineering, Hanyang University,
**School of Computer Engineering, Catholic University of Korea

요 약

많은 양의 비디오 데이터가 디지털화 되면서 사용자가 쉽게 자신이 원하는 비디오 데이터를 검색할 수 있는 내용 기반 검색이 필요하게 되었다. 내용 기반 검색을 위해서는 비디오 데이터를 연속된 카메라 모션으로 구성된 셋으로 나누고, 셋의 내용을 대표 할 수 있는 대표 프레임을 찾아야 한다. 대표 프레임은 비디오 데이터의 요약과 색인의 중요한 수단이다. 본 논문에서는 셋의 내용 기반으로 대표 프레임을 찾기 위해서 프레임에 존재하는 영역 정보를 바탕으로 셋의 내용을 알 수 있는 핵심 정보인 카메라 모션을 검출하고, 이를 기반으로 대표 프레임을 선택하는 방법을 제안한다.

1 서론

멀티미디어 사회가 되면서 많은 양의 비디오 데이터가 디지털화 되고 있다. 하지만 비디오 데이터는 일반 문서와는 달리 구조화(structuring)가 되어 있지 않아서, 양이 많아질수록 사용자가 원하는 정보를 신속하게 찾기 어렵다. 따라서 비디오 데이터를 효과적으로 검색할 수 있는 내용 기반 검색 시스템(content-based retrieval system)이 필요하게 되었다.

내용 기반 검색을 위해서는 비디오 데이터를 작은 단위로 나누고 관련된 정보를 색인하는 작업이 필요하다. 비디오 데이터는 크게 일관된 내용을 가지는 장면(scene)과 연속된 카메라 모션으로 구성된 셋(shot)으로 나눌 수 있다. 셋을 검출하는 방법은 영상의 픽셀 값이나 히스토그램의 차이를 이용하는 방법과 압축된 형태의 데이터에서 직접 얻을 수 있는 DCT 계수의 차이값을 이용하는 방법이 제시되었고, 이미 90% 이상 정확하게 검출되고 있다[1]. 또, 이러한 셋은 수직장의 프레임(frame)으로 구성이 되어 있고, 그 프레임들은 서로 많은 유사성을 가지고 있다. 따라서 셋의 내용을 대표하는 프레임을 찾을 수 있는데 이것을 대표 프레임(key-frame)이라고 한다. 대표 프레임은 수직장의 프레임을 한장의 프레임으로 표현하기 때문에 비디오 데이터 요약과 색인에 중요한 수단이다.

대표 프레임을 찾는 방법은 많이 연구가 되어 있으나 아직까지 셋의 내용을 기반으로한 일반적인 방법이 제시되지 못하고 있다. 본 논문에서는 셋의 내용을 알 수 있는 핵심 정보인 카메라 모션을 영역 정보를 바탕으로 검출하고, 이를 이용한 셋 내용 기반의 대표 프레임을 선택하는 방법을 제안한다.

2. 관련 연구

대표 프레임을 선택하는 방법은 크게 두 가지가 있다. 첫째는 미리 정해진 위치의 프레임을 선택하는 방법이고, 둘째는 프레임의 부가 정보를 이용해서 선택하는 방법이다.

첫번째 방법은 JACOB[2]에서 사용했다. 각 셋의 첫번째 프레임 또는 가운데 프레임을 대표 프레임으로 선택하는 방법이다. 이 방법의 장점은 간단하다는 점이지만 셋의 내용을 전혀 고려하지 않았다는 단점이 있다. 두번째 방법은 프레임의 색상이나 히스토그램 정보를 이용한 것과 움직임 정보를 이용한 것 등이 있다. Peter J. Macer[3]은 프레임의 색상 정보를 이용해서 대표 프레임을 선택했고, SMASH System[4]은 프레임의 히스토그램 차이를 이용해서 대표 프레임의 개수를 조정할 수 있는 방법을 제안 했다. Informecha™[5] 와 Wayne Wolf[6]은 셋내 프레임의 움직임 정보를 이용해서 대표 프레임을 찾는 방법을 제안했다. 옵티컬 플로우(optical flow)를 이용해서 프레임의 움직임 정보를 구한 후 움직임이 가장 적은 프레임을 대표 프레임으로 선택한다. 이 방법은 프레임이 정지해 있을 때를 대표 프레임으로 정하는 방법이기 때문에 셋 내에서 지속적인 움직임이 있는 경우는 대표 프레임을 찾기 힘든 단점이 있다.

3 카메라 모션 검출 및 대표 프레임 선택 방법

기존의 대표 프레임 선택 방법들은 셋의 내용을 기반으로 하지 못하고 있다. 따라서 본 논문에서는 셋의 내용을 알 수

있는 카메라 모션을 검출하는 방법을 제안하고, 이를 기반으로 대표 프레임을 선택할 수 있는 방법을 제안한다.

3.1 영역 정보를 이용한 카메라 모션의 검출 방법

카메라 모션에는 카메라의 위치가 고정되어 있으면서 수평 또는 수직 선회하는 팬(panning)과 틸트(tilting), 카메라의 위치가 좌우 이동하는 트래킹(tracking), 상하 이동하는 붐잉(booming), 전후 이동하는 달링(dollying), 그리고 확대 또는 축소하는 줌인(zoom in)과 줌아웃(zoom out)이 있다. 이러한 카메라 모션은 한 셋에 하나만 나타날 수도 있고 또는 동시에 여러 개가 혼합 될 수도 있다. 일반적으로 셋의 카메라 모션을 전부 찾아내는 것은 어려운 일이기 때문에 본 논문에서는 카메라의 위치가 고정되어 있고 동시에 하나의 동작만 있는 팬, 틸트, 줌인, 줌아웃의 경우만 고려한다.

셋의 카메라 모션을 검출하기 위해서는 먼저 셋내 프레임의 움직임 정보를 알아야 한다. MPEG 비디오 데이터에서는 모션 벡터(motion vector)를 직접 이용할 수 있지만 인코딩시 모션 벡터가 생략 될 수도 있기 때문에 본 논문에서는 움직임 플로우를 구해서 사용한다. 움직임 플로우를 구하는 방법은 여러가지가 있으나 상대적으로 어려움이 적은 Lucas 방법을 사용한다[7]. 움직임 플로우는 카메라 모션이 줌인의 경우 초점을 중심으로 발산하는 방향을 가지고, 줌아웃인 경우 수렴하는 방향을 가진다. 팬이나 틸트인 경우 움직임 플로우는 화면 전체적으로 일정한 방향을 가진다. 그림 1은 카메라 모션이 줌인과 팬인 프레임의 움직임 플로우를 보여준다. 하지만 실제 비디오 데이터는 움직이는 객체에 의해서 움직임 플로우가 그림 1과 같이 화면 전체적으로 일정하게 구해지지 않는 경우가 대부분이다. 따라서 영역을 찾아서 영역별로 움직임을 분석하는 것이 바람직하다.

프레임의 영역 정보를 찾기 위해서 본 논문에서는 MPEG 비디오 데이터로부터 먼저 DC 영상을 복원하고[8], 이 DC 영상으로부터 공간 및 시간적인 움직임성을 이용하여 비디오 데이터를 판별하는 단락으로 분할 한다. 관련된 단락으로 분할하기 위해서, 각각의 영상으로부터 중요한 객체가 차지하는 주된 영역(dominant region)들을 찾고, 시간 축을 따라 각각의 DC 영상에서 주된 영역의 대응 관계(correspondence)를 찾아 각 영역들의 시간축에 따른 변화 정도를 이용하여 비디오 데이터에 존재하는 영역들을 찾아낸다[9].

영역이 찾아지면 분할된 영역을 기반으로 움직임의 주방향(dominant direction, DD)과 움직임 세기(motion intensity, MI)를 계산해서 줌을 검사한 후 팬(틸트)을 검사한다. 줌을 검사하기 위해 먼저 초점(FOE or FOC)을 계산한다. 초점은 모든 플로우 벡터의 방향으로 선을 그려서 가장 많은 선이 교차하는 김으로 정한다. 초점이 정해지면 그림 2의 a)와 같이 초점을 기준으로 수직, 수평방향의 초점축을 그려서 영상을 네개의 윈도우(window)로 나누고, 각 윈도우에 속한 영역 부분의 플로우 벡터를 이용해서 윈도우의 주방향과 움직임 세기를 계산한다. 주방향은 플로우 벡터중 개수가 가장 많은 방향이고, 움직임 세기는 플로우 벡터 크기 총합을 플로우 벡터 개수로 나눈 값이다. 주방향과 움직임 세기가 구해지면 움직임 세기가 임계값 이상인 윈도우의 주방향을 검사해서 주방향이 초점축을 향하여 발산하면 줌인이고, 수렴하면 줌아웃이다.

줌이 검출안되면 팬(틸트)을 검사한다. 팬(틸트)은 먼저 각 영역별로 주방향과 움직임 세기를 구한다. 팬(틸트)의 경우는 보통 하나의 영역이 하나의 주된 방향으로 움직이기 때문에 주방향과 움직임 세기를 동일해하는 다른 방법으로 구한다. 주방향은 영역안의 플로우 벡터중 가장 많은 방향의 개수가

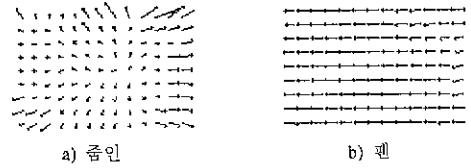


그림 1 카메라 모션에 따른 움직임 플로우

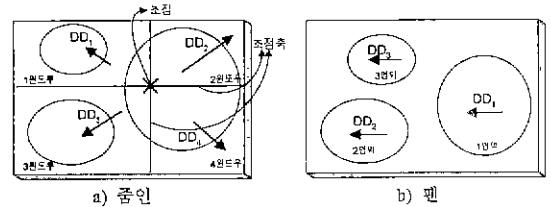


그림 2 영역의 주 방향과 카메라 모션

영역내 플로우 벡터의 총 개수의 1/2 보다 많으면 그 방향을 주방향으로 정한다. 움직임 세기는 영역내의 플로우 벡터 중에서 방향이 주방향과 같은 벡터 크기의 총합을 영역내 플로우 벡터의 개수로 나눈 값이다. 식으로 나타내면 아래와 같다

$$DD_r = \begin{cases} \theta_r & \text{if } N_{r, \theta} \geq N_{r, \theta} / 2 \\ \text{don't care} & \text{otherwise} \end{cases}$$

$$MI_r = \sum_{x,y} m(x,y) \times \psi(\alpha(x,y), DD_r) / N_{r, \theta} \quad (x,y \in r \text{ region})$$

$$N_{r, \theta} = \text{MAX} \{ PH_r(\theta_r) \} \quad (\theta_r = 0, \theta_r = 45^\circ, \theta_r = 315^\circ)$$

$$\psi(\alpha, \theta) = \begin{cases} 1 & \text{if } \alpha = \theta \\ 0 & \text{otherwise} \end{cases}$$

DD_r, MI_r 은 r 영역의 주 방향과 움직임 세기이고, PH_r 은 r 영역내 각 방향 플로우 벡터의 개수를 가지고 있는 방향 히스토그램(phase histogram)이다. $m(x,y), \alpha(x,y)$ 는 x,y 좌표에 있는 플로우 벡터의 크기와 각도이다. $N_{r, \theta}$ 은 r 영역안에 있는 플로우 벡터의 총 개수이다. 각 영역의 주방향과 움직임 세기가 정해지면 움직임 세기가 임계값 이상인 영역의 주방향을 검사해서 주방향이 일정한 방향이면 팬(틸트)로 검출한다. 줌과 팬(틸트)은 것도 검출이 안되면 다른 카메라 모션이 있는지를 검사한다. 팬(틸트)에서 구한 영역의 움직임 세기가 임계값 이상인 영역이 하나라도 있으면 다른 카메라 모션이 있다고 검출하고, 하나도 없으면 카메라 모션이 없다고 검출한다.

셋내 프레임들의 카메라 모션이 찾아지면 셋의 카메라 모션을 결정한다. 셋내 프레임 개수에서 카메라 모션이 없다고 검출된 프레임 개수를 팬 값의 1/2 이상이 하나의 카메라 모션으로 검출되면 그것으로 셋의 카메라 모션을 정한다.

3.2 대표 프레임 선택 방법

셋의 카메라 모션이 정해지면 제작자의 의도를 알 수 있기 때문에 이를 이용해 셋 내용 기반의 대표 프레임을 선택 할 수 있다. 카메라 모션에 따른 제작자의 의도는 표 1에 정리했다. [10] 팬은 보통 주변의 풍경을 묘사하거나, 서로 떨어진 두 장소를 연결 시키기 위해 사용된다. 따라서 처음과 마지막 영상을 대표 프레임으로 선택 함으로써 제작자가 보여주고자 하는 풍경이나 장소를 나타낼 수 있다. 줌인 경우는 제작자가

주인공의 감정이나 현재 상황을 강조하기 위해서 사용된다 따라서 가장 많이 확대된 마지막 영상을 대표 프레임으로 선택한다 줌아웃의 경우도 줌인과 마찬가지로 갑자기 시청자의 시각을 넓히 주면서 상황을 강조하는데 사용되기 때문에 가장 강조된 마지막 영상이 대표 프레임이 된다 카메라 모션이 찾아지지 않는 경우에는 간단한 방법으로 셋의 첫번째 프레임을 대표 프레임으로 선택한다

표 1 카메라 모션에 따른 제작자 의도와 대표 프레임

카메라 모션	제작자 의도	대표 프레임
펜 킬트	주위 풍경을 보여준다 멀어진 장소를 연결시킨다	처음과 마지막 프레임
줌인	심황이나 감정을 강조한다	마지막 프레임
줌아웃	시각을 넓게 보여 주면서 심황을 강조한다	마지막 프레임
검출 실패	알수없다	첫번째 프레임

4 실험결과

실험에는 펜, 줌인, 줌아웃의 카메라 모션이 있는 MPEG 비디오 데이터를 사용했다 검출의 효율을 위해 플로우 벡터의 방향을 8개로 양자화했고, 플로우 벡터는 크기가 임계값 이상인 것만 고려 했다. 그림 3의 a)는 카메라 동작이 펜인 flower.mpg 이다 1번에서 50번 프레임의 카메라 모션 검출 결과 50장의 영상중에 펜으로 검출된 프레임은 32개이고 잘못 검출된 프레임은 14개였다. 카메라 모션이 없다고 검출된 프레임수는 4장인데 이것은 옵티컬 플로우를 구할 수 없는 처음 두장과 마지막 두장의 프레임이다. 따라서 셋의 카메라 모션은 펜으로 검출되었고 대표 프레임은 1번과 50번 프레임으로 선택 되었다 b)는 카메라 모션이 줌인인 wg_cs_5.mpg 의 85번에서 130번 프레임이다 검출 결과 46장의 프레임 중에서 줌인은 27장, 검출 실패는 7장, 모션 없음은 12장이 나와서 셋의 카메라 모션은 줌인으로 검출되었고 대표 프레임은 130번 프레임으로 선택 되었다. c)는 카메라 모션이 줌아웃인 wg_cs_5.mpg 의 151번에서 203번 프레임이다 검출 결과 53장의 프레임 중에서 줌아웃이 20장, 잘못된 검출이 2장, 모션 없음은 21장이 나와서 셋의 카메라 모션은 줌아웃으로 검출되었고 대표 프레임은 203번 프레임이 선택 되었다.

5. 결론

본 논문에서는 비디오 데이터의 영역 정보와 옵티컬 플로우를 이용해 셋의 카메라 모션을 검출하였고, 이를 기반으로 대표 프레임을 선택했다.

프레임을 영역으로 분할해서 영역내의 옵티컬 플로우만 사용해서 움직이는 객체에 의해 카메라 모션이 잘못 검출되는 것을 방지할 수 있었다 카메라 모션을 이용해서 대표 프레임을 선택했기 때문에 셋 내용 기반으로 대표 프레임을 찾을 수 있었다.

앞으로, 칼라 및 텍스트어(lexic) 정보들을 결합하여 보다 강건한 내용 기반의 비디오 처리 시스템을 구현할 수 있을 것으로 기대된다

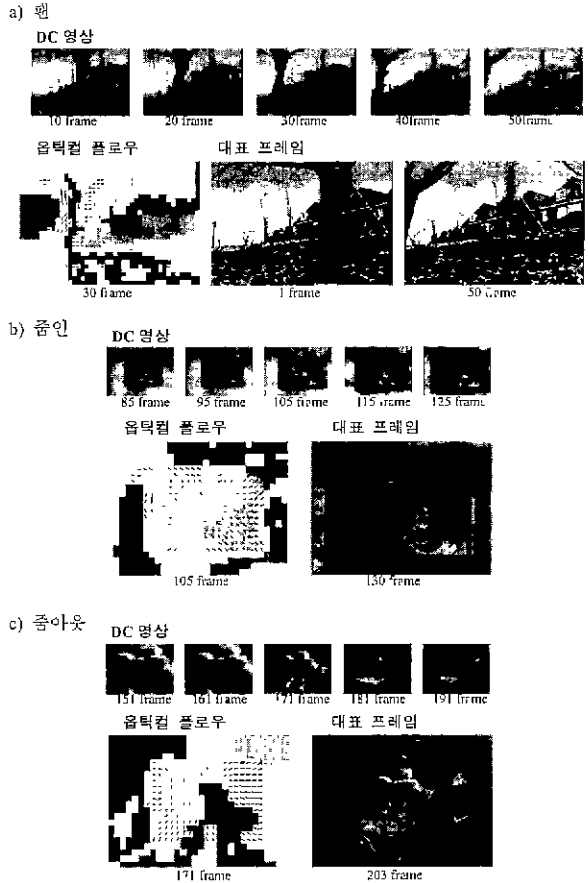


그림 3 카메라 모션과 찾아진 대표 프레임

참고문헌

- [1] Grosky, Jain, Mehrotra, *The Handbook of Multimedia Information Management*, Prentice Hall, 1997
- [2] Marco La Cascia, Edoardo Arduzzone, "JACOB: Just A Content-Based Query System for Video Databases", *IEEE ICASSP'96*, 1996.
- [3] Peter J. Macer, Peter J Thomas, Nuhman Chalabi, Jhon F Meech, "Finding the Cut of the Wrong Trousers: Fast Video Search Using Automatic Storyboard Generation", *CHI'96*, 1996
- [4] R.L. Lagendijk, A. Hanjalic, M. Ceccarelli, M. Solovic, E. Persoon, "Visual Search in a SMASH System", *IEEE ICIP'96*, 1996
- [5] A. G Hauptmann, M. A. Smith, "Text, Speech, and Vision for Video Segmentation The Informedia™ Project", *AAAI'95*, 1995
- [6] Wayne Wolf, "Key Frame Selection by Motion Analysis", *IEEE ICASSP'96*, 1996.
- [7] J.L. Barron, S.S. Beauchemin, D.J. Fleet, "On Optical Flow", *AIICSR*, 1994.
- [8] B.L. Yeo, B. Liu, "On the extraction of DC sequences from MPEG compressed video", *IEEE ICIP'95*, 1995.
- [9] H.B. Kang, "A New 3-D Segmentation Method for Region-Based Video Coding", *IEEE ICASSP'96*, 1996.
- [10] Steven d Katz, *영화연출론 Shot by Shot*, (주)시공사, 1998.