

평균값을 이용한 통합형 카메라 동작 추출 시스템 개발

한규서, 이재연, 정세윤, 배영래
한국전자통신연구원 영상정보처리연구원

e-mail : kyuseo@etri.re.kr

A Hybrid System for Extracting Camera Motion from Average Projected Image

Kyu-Seo Han, Jae-Yeon Lee, Se-Yoon Jeong, Young-Lae Bae
Electronics and Telecommunication Research Institute

요 약

동영상 검색 및 관리 기술의 필요성 증가에 따라 동영상 검색 기법들이 제안되고 있으며, 특히 내용 기반 검색 기법에 대두되고 있다. 본 논문에서는 내용 기반 요소중의 하나인 카메라 동작 검출을 위한 시스템을 제안한다. 본 시스템은 영상의 평균값을 이용하여 제작된 투영 영상상에서 추출된 edge의 방향 성분으로 구성된 optical flow에 따라 적절한 카메라 동작을 검출한다. 효과적인 카메라 동작 요소 검출을 위하여 평균값으로 구성된 투영 영상을 검출하고자 하는 카메라 동작 요소에 따라 달리 제작한다. 투영 영상은 전체적인 영상에 대하여 제작하는 방법과 sub-block으로 나눈 영상에 따라 각 블록별로 제작하는 방법으로 나누어진다.

1. 서 론

멀티미디어 데이터의 증가와 함께 효과적인 동영상 검색 및 관리 기술의 필요성이 증가하고 있다. 특히 사용자의 요구에 맞는 동영상 검색의 필요성이 대두되고 있는 가운데 효과적인 검색을 위하여 동영상의 분류 및 관리에 있어 동영상에 포함된 내용을 기반으로 하는 기법의 연구가 활발히 진행되고 있는 가운데, 국제적으로는 MPEG-7과 같이 동영상 검색 및 분류를 위한 내용 요소(content descriptors)들에 대한 정의가 표준화 되어지고 있다. 이러한 요소들 중에서 영상내의 카메라 동작에 관한 요소도 지정되어 있다. 또한 CG와 실사 영상의 효과적인 합성을 위하여 일반 실사 영상내에서의 카메라 동작 요소의 추출이 중요한 작업으로 알려져 있다.

카메라 동작 요소로는 Tilt, Zoom, Panning, Dollying 등이 있으며 추출 방법으로는 Optical Flow를 이용하는 방법이 주로 제안되어지고 있다[1]. 그러나 Optical

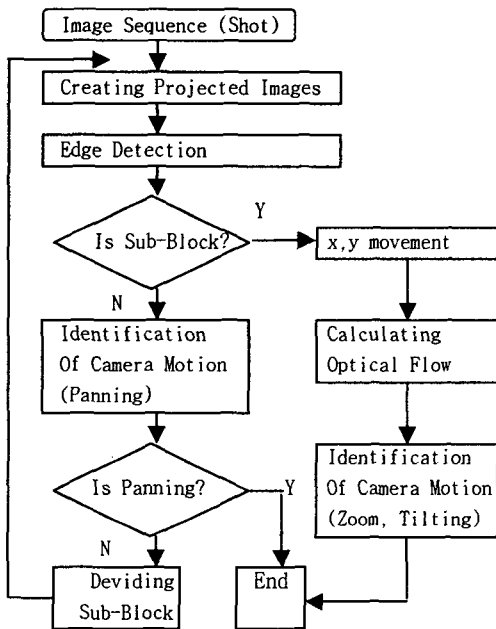
Flow를 이용하는 방법은 과도한 계산량을 요구하므로 실시간 처리가 어려운 상태이다. 또한 각 frame의 line당 평균값을 이용한 방법[2]도 제안되었는데 edge추출을 위한 level양자화에 민감하게 반응하여 카메라 동작 추출에 어려움이 있다. 이를 개선하기 위하여 본 논문에서는 frame전체와 sub-blockframe상의 수직, 수평 방향의 line당 평균값으로 구성된 투영 영상(projected image)를 이용한 카메라 동작 요소 검출 방법을 제안한다. 제안하는 방법은 기존의 방법[2]이 frame전체에 대한 평균값을 이용하여 발생할 수 있는 정보 손실을 최소화하고 또한 영상의 각 sub-region의 정보를 이용함으로써 좀더 효과적인 카메라 동작 요소 추출을 가능하게 만든다. 우선 기존의 방법[2]을 이용하여 panning을 검출한 뒤 그 외의 카메라 동작은 각 frame영상을 일정한 크기의 sub-block을 분리하여 각 block당 수평, 수직 방향의 평균값을 가지는

2 개의 투영 영상을 얻는다. 이렇게 얻은 2 개의 투영 영상내에서 edge 방향을 추출하여 얻은 block 내의 이동 속도를 카메라 동작 파라미터, 이동 속도로 표현되는 비선형 방정식에 대입하여 카메라 동작 요소를 검출한다. 제안한 방법에서는 기존의 방법[2]에서 추출이 어려운 zoom-in/out, tilt-up/down 이 구별되어 보다 정밀한 카메라 동작 요소를 검출할 수 있다.

다음 장에서 제안한 방법을 설명하며 3 장에서는 제안한 방법에 따른 실험과 결과를 설명한다. 4 장에서는 결과에 따른 고찰로서 본 논문은 구성되어 있다.

2. 투영 영상을 이용한 카메라 동작 검출

제안하는 방법은 크게 두단계로 나뉘어진다. 첫번째 단계로 기존의 방법[2]을 이용하여 panning 동작을 검출한 뒤 두번째 단계로 sub-block 당 투영 영상을 이용하여 zoom, tilt 동작을 검출한다. 단계별로 구성된 순서도를 그림 1.에 나타낸다.



<그림 1.> 제안하는 방법의 단계별 순서도

M x N 크기를 가지는 T 개의 frame 으로 구성된 동영상상에서의 각 frame 별로 식 (1)을 이용하여 수직 방향의 투영 영상을 제작한다. 제작된 투영 영상내에 sobel operator[3]을 이용하여 각 edge 의 방향 성분을 구하여 이에 따라 카메라 동작 중 panning 을 구별한다.

$$P^v(t, k, l) = \frac{1}{N} \sum_{i=1}^N I(t, k, i) \quad (t = 1, 2, 3, \dots, T) \quad (1)$$

edge 의 방향 성분에 따라 panning 의 종류가 결정된다. 시계 방향으로 0° ~ 90° 이면 pan right, 0° ~ -90° 이면 pan left 로 결정된다.

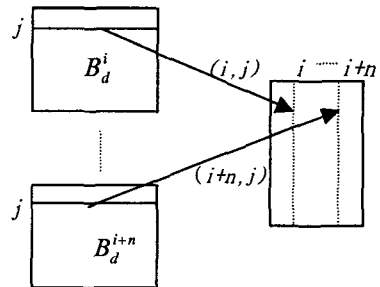
한편 식 (1)에서 구한 투영 영상내의 edge 성분이 설정한 임계값 이하로 나오게 되면 sub-block 으로 frame 을 나누어 투영 영상을 제작한다.

각 sub-block 에 대한 수평 방향과 수직 방향의 투영 영상은 식 (2)와 같다. 각 sub-block 은 m x n 의 크기를 가지며 총 D 개로 구성된다.

$$P_d^h(k, l) = \frac{1}{m} \sum_{i=1}^m B_d^h(i, l) \quad (k = 1, 2, 3, \dots, T) \quad (2)$$

$$P_d^v(k, l) = \frac{1}{n} \sum_{j=1}^n B_d^v(k, j) \quad (l = 1, 2, 3, \dots, T)$$

여기서, k, l 은 frame 순서 및 투영 영상내의 pixel 의 좌표를 나타내며 B_d는 d 번째 sub-block 를 의미하며 P_d^h, P_d^v는 d 번째 sub-block 의 수평, 수직방향의 투영 영상을 나타낸다. 식(2)에 따르는 sub-block 당 수평, 수직방향의 투영 영상중 수평 방향 투영 영상 P_d^h 의 구조는 그림 1.과 같다.



<그림 2.> sub-block 당 수평 투영 영상의 구조

위의 식(2)로 구해진 각 sub-block 에 대한 수평 투영 영상 P_d^h 와 수직 투영 영상 P_d^v 상에 존재하는 edge 들의 방향성분을 구한다. 본 논문에서는 sobel operator[3]를 사용하여 각 edge 들의 방향 성분을 구하였다. 이렇게 구해진 방향 성분은 시간에 따른 이동 성분만을 가지고 있으나 실제적으로 원영상내에서 x, y 방향의 공간적인 이동 성분을 추출하는 것이 필요하다. edge 방향 성분은 투영 영상의 시간을 대표하는 축의 위치에 따라 제한된다. 수평 투영 영상

P_d^h 의 경우 edge의 방향은 $\pm 90^\circ$ 로 제한되며 수직 투영 영상 P_d^v 의 경우 방향은 $180^\circ \sim 360^\circ$ 로 제한된다. 각 sub-block의 x, y 방향의 이동 성분은 식(3)과 같이 구한다.

$$\begin{aligned} x_d &= T / \tan \theta_{v,d} \\ y_d &= T * \tan \theta_{h,d} \end{aligned} \quad (3)$$

여기서 x_d, y_d 는 각 sub-block 당 x, y 방향의 이동 성분이며 T 는 총 frame 수, $\theta_{v,d}, \theta_{h,d}$ 는 sub-block 당 투영 영상내의 edge 방향을 나타낸다. 식(3)에서 얻어진 x, y 방향의 이동 성분과 영상의 총 frame 수 T 를 이용하여 sub-block 당 이동 속도를 구할 수 있다. 한편 카메라 동작은 식(4)와 같은 방정식으로 나타낼 수 있다

$$\begin{aligned} u_x &= \frac{xy}{f} R_x - f(1 + \frac{x^2}{f^2}) R_y + f \tan^{-1}(\frac{x}{f}) \cdot (1 + \frac{x^2}{f^2}) R_{zoom} \\ u_y &= -\frac{xy}{f} R_y - f(1 + \frac{y^2}{f^2}) R_x + f \tan^{-1}(\frac{y}{f}) \cdot (1 + \frac{y^2}{f^2}) R_{zoom} \end{aligned} \quad (4)$$

여기서, u_x, u_y 는 (x, y) 위치상에서의 이동 속도, R_x 는 Tilting, R_y 는 Panning, R_{zoom} 은 Zooming, f 는 카메라의 focal length를 나타낸다[4].

식(4)에 의하여 카메라 동작을 추출하기 위해서는 각 카메라 동작 요소 파라미터들은 Pseudo Inverse를 이용하여 계산되어진다. 식(4)에 각 sub-block 당 이동 속도를 대입하여 카메라 동작 요소를 추출한다. 이때 x, y 위치값은 각 sub-block의 중앙값으로 한다.

각 sub-block 당 추출되어진 카메라 동작 요소값들을 수집하여 가장 많이 추출되어진 카메라 동작 요소값을 영상내의 카메라 동작 요소로 결정한다.

3. 실험 및 결과

본 논문에서 제한한 방법의 검증을 위하여 총 55개의 영상으로 이루어진 Ground-Truth를 작성하였다. 작성된 Ground-Truth의 구성은 표 1.과 같다. 실험에 사용된 영상은 352 x 288의 크기를 가지며 각 sub-block의 크기는 16 x 16으로 하였다. 우선 식(1)에 의하여 수직 방향의 투영 영상을 구하여 영상내의 edge 방향 성분을 구하여 panning 동작을 검출하였다. Egdg 방향 성분이 임계값을 초과하지 못하였을 경우 식(2)에 따라 각 sub-block 당 수평, 수직 방향의 투영 영상을 구한다.

<표 1.> Ground-Truth의 구성

| Camera operation | Number |
|------------------|--------|
| Tilt(up/down) | 9 |
| Pan(left/right) | 17 |
| Zoom(in/out) | 23 |
| No operation | 6 |

실험상에서 임계값은 전체 frame 수의 75%로 정하였다. 식(2)에 따라 가로, 세로 각 396개의 투영 영상을 만들고 이로부터 각 sub-block 당 edge 방향 성분을 구한다. 구해진 edge 방향 성분을 식(3)에 대입하여 x, y 좌표상의 이동 방향과 크기를 식(4)에 대입함으로써 각 sub-block 당 카메라 동작 요소 성분을 검출한다. 검출 카메라 동작 요소는 총 8개이며 카메라 동작 요소가 없는 영상도 포함시켰다. Ground-truth에 따른 실험 결과는 표 2.와 같다.

<표 2.> 제안한 방법에 의한 카메라 동작 추출 결과

| Camera operation | Number (correct/total) | Correct(%) | Avg. time(s) |
|------------------|------------------------|------------|--------------|
| Tilt(up/down) | 7/9 | 77.7 | 2.22 |
| Pan(left/right) | 15/17 | 88.2 | 3.48 |
| Zoom(in/out) | 21/23 | 91.3 | 2.37 |
| No operation | 2/6 | 33.3 | 1.85 |

실험 결과에서 알 수 있듯이 전체적으로 카메라 동작이 없는 영상의 경우를 제외하고 85% 이상의 카메라 동작을 검출할 수 있었다. 기존의 방법[5]와 비교해서 panning의 경우 10% 이상의 검출 향상도를 보였다. 이는 기존의 방법[5]에서 sub-block 당 카메라 동작 요소를 추출할 때 카메라 동작 여부와는 관계없는 영상내의 강한 edge 성분이 이동 성분으로써 작용하는 것을 제거할 수 있었던 것으로 판별된다. 특히 panning의 경우 각 sub-block 상에서 발생하던 공간적 분포에 따른 edge 성분을 제거하여 효과적인 카메라 동작 검출을 이룰 수 있었다. 즉, sub-block으로의 분리에 의해 나타났던 공간 정보를 가지고 있는 edge 성분이 사라지게 되었다.

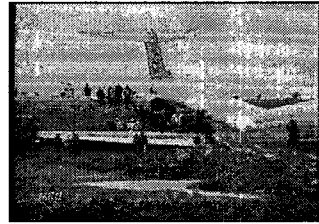
그림 3.에서는 panning의 경우에 나타나는 투영 영상과 edge 성분을 추출한 것을 보여주고 있다. 또한 그림 4.에서는 제안한 방법을 Zoom-in 요소가 있는 영상에 적용하였을 때 얻어진 Optical flow를 나타내고 있다. 그림 3.(a)는 원본 영상의 일부를 나타내고 있으며, (b), (c)는 투영 영상상에서의 시간에 따른 x, y 이동 방향 성분을 나타내고 있다. 그림 3.(d)에서 (b), (c)를 이용해 얻은 각 sub-block 상의 optical flow를 보여주고 있다.

4. 결론

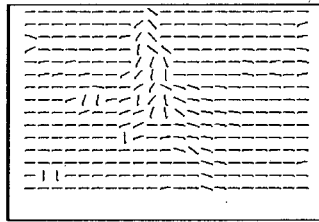
이상과 같이 영상의 평균값을 이용한 투영 영상을 이용하여 카메라 동작 검출 방법에 대하여 설명하였다. 실험결과에서 나타난 바와 같이 panning 동작의 경우 개선된 edge 성분 판별에 의해 검출이 가능하였고 zoom, tilting의 경우 제안된 sub-block 상의 투영 영상 상에서 추출된 optical flow를 이용하여 검출이 가능하였다. 제안한 방법으로 8개의 카메라 동작에 대한 검출이 가능하였으며 그 정확도 또한 90%에 근접하였다. 향후 연구에서는 실시간 처리와 실험 데이터의 확대를 통하여 일반화 시켜야 할 것이다.

참고 문헌

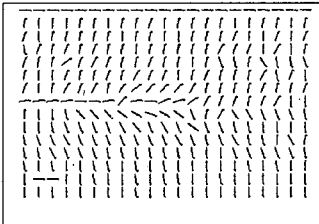
[1]J.L.Barron, et.all, " Systems and Experiment:Performance. of Optical Flow Techniques" Int' l J. of Computer Vision, Vol. 12, 1994 pp.43-77
 [2]P.Jolly, H.K.Kim, " Efficient automatic analysis of camera work and microsegmentation of video using spatiotemporal images" , Image Commu. Vol. 8, 1996, pp.295-307
 [3]M. A. Sid-Ahmed, " Image Processing" , McGraw-Hill, 1994
 [4]ISO/IEC MPEG-7 Visual Part of XM, Version 8.0 pp 56-58
 [5]한규서 외 3인, " Sub-block 투영 영상을 이용한 카메라 동작 검출 방법" . 정보과학회 추계학술대회 2001, 2001.10.



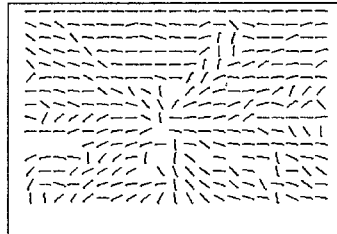
(a) 원본 영상



(b) 세로 투영 영상 P_d^h 에 대한 이동 성분



(c)가로 투영 영상 P_d^v 에 대한 이동 성분

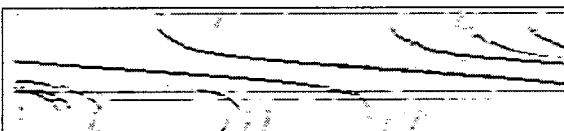


(d)각 sub-block 당 optical flow

<그림 4.> Optical Flow in Zoom-in



(a) 세로 투영 영상(pan left)



(b) edge 검출

<그림 3.> 카메라 동작(panning) 검출