평균값을 이용한 통합형 카메라 동작 추출 시스템 개발

한규서, 이재연, 정세윤, 배영래
한국전자통신연구원 영상정보처리연구팀

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-Laee 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 요소중 하나인 카메라 동작 검출을 위한 시스템을 제안한다. 본 시스템은 영상의 평균값을 이용하여 제작된 투영 영상에서 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 단일 제작한다. 투영 영상은 전체적인 영상에 대하여 제작하는 방법과 sub-block 으로 나눈 영상에 따라 각 별도로 제작하는 방법으로 나누어진다.

1. 서 론


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

2. 투영 영상을 이용한 카메라 동작 검출


![Image Sequence (Shot)]

Creating Projected Images

Edge Detection

Is Sub-Block?

Y

x, y movement

N

Identification Of Camera Motion (Panning)

Calculating Optical Flow

Is Panning?

Y

N

Deviding Sub-Block

End

프레임에 대한 sub-block 당 수평 투영 영상의 구조

위의 식(2)로 구해진 각 sub-block에 대한 수평 투영 영상 \( P_d^k \)와 수직 투영 영상 \( P^v_d \) 상에 존재하는 edge 등의 방향성분을 구한다. 본 논문에서는 sobel operator[3]를 사용하여 각 edge 등의 방향성분을 구하였고, 이렇게 구해진 방향성분은 시간에 따른 이동 성분만을 가지고 있으므로 실제적으로 원영상에서 \( x, y \) 방향의 공간적인 이동 성분을 추출하는 것이 필요하다. edge 방향 성분은 투영 영상의 시간을 대표 하는 측의 위치에 따라 제한된다. 수평 투영 영상

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

edge의 방향 성분에 따라 panning의 종류가 결정되며. 시계 방향으로 0° ~ 90°이면 pan right, 0° ~ 90°이면 pan left로 결정된다.

한편 식 (1)에서 구한 투영 영상내의 edge 성분이 설정한 임계값 이하로 나오게 되면 sub-block으로 frame을 나누어 투영 영상을 제작한다.

각 sub-block에 대한 수평 방향과 수직 방향의 투영 영상은 식 (2)와 같다. 각 sub-block은 \( m \times n \) 크기로 나누어 총 4개로 구성된다.

\[
P^h_d(k,l) = \frac{1}{m} \sum_{i=1}^{m} B^h_d(i,l) \quad (k=1,2,3,...,T)
\]

\[
P^v_d(k,l) = \frac{1}{n} \sum_{j=1}^{n} B^v_d(k,j) \quad (l=1,2,3,...,T)
\]

여기서, \( k, l \) 은 frame 순서 및 투영 영상의 pixel의 좌표를 나타내며 \( B^h_d \)는 \( d \) 번째 sub_block을 의미하며 \( P^h_d \). \( P^v_d \)는 \( d \) 번째 sub_block의 수평, 수직방향의 투영 영상을 나타낸다. 식(2)에 따르면 sub-block 당 수평, 수직방향의 투영 영상 수평 방향 투영 영상 \( P^h_d \)의 구조는 그림 1과 같다.

<그림 2> sub-block 당 수평 투영 영상의 구조
\[ P_d^b \]의 경우 \textit{edge}의 방향은 ±90\degree로 제한되며 수직 투영 영상 \( P_d^v \)의 경우 방향은 180\degree~360\degree로 제한된다. 각 sub-block의 \( x,y \) 방향의 이동 성분은 식(3)과 같이 구한다.

\[
x_d = T / \tan \theta_{vd} \\
y_d = T \cdot \tan \theta_{hd}
\]  
(3)

여기서 \( x_d, y_d \)는 각 sub-block당 \( x,y \)방향의 이동 성분이며 \( T \)는 종 frame 수, \( \theta_{vd}, \theta_{hd} \)는 sub-block당 투영 영상내의 edge 방향을 나타낸다. 식(3)에서 얻어진 \( x,y \)방향의 이동 성분과 영상의 종 frame 수 \( T \)를 이용하여 sub-block당 이동 속도를 구할 수 있다. 한편 카메라 동작은 식(4)과 같은 방정식으로 나타낼 수 있다.

\[
u_x = \frac{\partial y}{\partial f} R_s - f(1 + \frac{\partial^2}{\partial f^2}) R_s + f \tan^{-1}(\frac{\partial}{\partial f}) R_{zoom} \]  
(4)

\[
u_y = -\frac{\partial x}{\partial f} R_s - f(1 + \frac{\partial^2}{\partial f^2}) R_s + f \tan^{-1}(\frac{\partial}{\partial f}) R_{zoom}
\]

여기서, \( \nu_x, \nu_y \)는 \((x,y)\) 위치상에서의 이동 속도, \( R_s \)는 Tilting, \( R_s \)는 Panning, \( R_{zoom} \)은 Zooming, \( f \)는 카메라의 focal length을 나타낸다[4].

식(4)에 의하여 카메라 동작을 추적하기 위해서는 각 카메라 동작 요소 파라미터들은 Pseudo Inverse를 이용하여 계산되게 된다. 식(4)에 각 sub-block당 이동 속도를 대입하여 카메라 동작 요소를 추출한다. 이때 \( x,y \) 위치값은 각 sub-block의 중앙값으로 한다.

각 sub-block당 추출된 카메라 동작 요소 값을 수집하여 가장 많이 추출된 카메라 동작 요소값을 영상내의 카메라 동작 요소로 결정한다.

3. 실험 및 결과

본 논문에서 제안한 방법의 검증을 위하여 총 55개의 영상으로 이루어진 Ground-Truth를 작성하였다. 작성된 Ground-Truth의 구성은 표 1과 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 영상 내의 edge 방향 성분을 구하여 panning 동작을 검출하였다. Egdg 방향 성분의 임계값을 초과하지 못하였을 경우 식(2)에 따라 각 sub-block당 수평, 수직 방향의 투영 영상을 구한다.

<table>
<thead>
<tr>
<th>Camera operation</th>
<th>Ground-Truth의 구성</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>Tilt(up/down)</td>
<td>9</td>
</tr>
<tr>
<td>Pan(left/right)</td>
<td>17</td>
</tr>
<tr>
<td>Zoom(in/out)</td>
<td>23</td>
</tr>
<tr>
<td>No operation</td>
<td>6</td>
</tr>
</tbody>
</table>

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

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


그림 3에서는 panning의 경우 나타난 투영 영상과 edge 성분을 추출한 것을 보여주고 있다. 또한 그림 4에서는 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 영상에서의 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 시키려 할 것이다.

참고 문헌


<그림 3.> 카메라 동작(panning) 검출

<그림 4.> Optical Flow in Zoom-in