일반화된 Hough 변환기법을 이용한 MPEG2 압축영역에서의 카메라의 움직임 해석

유 원 영[†] · 최 정 일[††] · 이 준 환[†††]

요 약

본 논문에서는 MPEG2 비디오 스트림에서 완전한 복호화 과정없이 직접 얻을 수 있는 압축 정보들을 이용하여 간단하고 효과적인 카메라 움직임(Camera Operation) 정보 추출방법을 제안하였다. 제안된 방법은 순서적이지 못한 움직임 벡터를 갖는 MPEG2 비디오 스트림에서, 선지식으로 알고있는 예측 프레임들의 속성을 이용하여 움직임 벡터(motion Vector)로부터 광 플로우(optical flow)를 추출하였으며, 일반화된 Hough 변환기법 이용하여 카메라 움직임 모델의 기본 요소인 팬(Pan), 틸트(Tilt), 줌(Zoom) 등을 근사적으로 추정하였다. 제안된 방법은 농구경기 및 축구경기 비디오에 대해 적용하여 기존의 최소자승방법(Least Mean Square) 방법보다 양호한 실험결과를 얻을 수 있었다.

제안된 카메라 움직임 추정은 비디오 스트림에서 작접 정보를 얻음으로써 계산속도의 향상은 물론, 카메라의 움직임을 활용하는 내용기반의 비디오 검색 및 분석에서 유용한 요소 기술이 되리라고 기대된다. 또한 MPEG2 비디오 스트림에서 이러한 개발 기술들은 압축영상을 복호화 하지않고 객체들의 움직임만을 탐색하거나 추적하는데 활용될 것으로 예상된다.

Analysis of Camera Operation in MPEG2 Compressed Domain Using Generalized Hough Transform Technique

Won-Young Yoo[†] · Jeong-Il Choi[††] · Joonwhoan Lee[†††]

ABSTRACT

In this paper, we propose a simple and efficient method to estimate the camera operation by using compressed information, which is extracted directly from MPEG2 stream without complete decoding. In the method, the motion vector is converted into approximate optical flow by using the feature of predicted frame, because the motion vector in MPEG2 video stream is not regular sequence. And they are used to estimate the camera operation, which consist of pan, tilt, and zoom by Hough transform technique. The method provided better results than the least square method for video stream of basketball and soccer games.

The proposed method can have a reduced computational complexity because the information is directly obtained in compressed domain. Additionally it can be a useful technology in content-based searching and analysis of video information. Also, the estimated camera operation is applicable in searching or tracking objects in MPEG2 video stream without decoding.

1. 서 론

인터넷을 통한 멀티미디어의 활용이 점점 증가되고 있는 시점에서 대용량의 비디오 정보의 효율적인 분석과 검색의 필요성이 점차 증대되고 있다. MPEG 4와 7에서처럼 객체기반의 압축에서는 압축된 비트 스트림에 객체 및 배경에 대한 정보를 명시적으로 담고 있기 때문에 이들을 직접 접근하기가 용이하나, 자연영상의 경우는 아직 활용이 보편화되지는 않고 있으며, 더지

† 준 회 원 : 전북대학교 대학원 전자공학과
†† 준 회 원 : 전북대학교 대학원 전자공학과
††† 정 회 원 : 전북대학교 전자공학과 교수
논문접수 : 2000년 7월 24일, 심사완료 : 2000년 11월 24일

털 TV나 많은 응용 프로그램의 저장물 등은 MPEG1과 MPEG2로 저장되어지고 사용되고 있다. 이러한 이유 때문에 아직도 프레임기반의 압축물을 검색하고 분류하기 위한 노력이 계속되고 있고, 빠른 처리속도와 효율의 향상을 위하여 압축정보의 활용을 시도하고 있다. 특히 DCT계수와 같은 매크로 블록(Macro Block)의 정보는 장면전환 인식에 사용되고 있고, 움직임 벡터(Motion Vector)와 같은 정보는 카메라 움직임(Camera Operation) 또는 객체인식 및 추적분야에 활용되고 있는 추세이다[1-5].

본 논문에서는 MPEG2의 완전한 복호화 과정없이 MPEG2의 비디오 스트림이 포함하고 있는 움직임 벡터들을 이용하여 카메라의 움직임을 추정하려고 한다. 그러나 MPEG2의 움직임 벡터들은 압축의 효율을 높이기 위해서 구한 것이기 때문에 I 프레임에서는 움직임 정보에 해당하는 움직임 벡터가 없고, B프레임에서는 앞 프레임과 뒤 프레임에서의 예측한 움직임 벡터를 가지고 부호화하기 때문에 순서적인 예측 흐름인 광 플로우를 가지지 못한다. 또한 MPEG은 압축효율의 향상을 위해 공간적인 면적상관 방법(area-correlation method)을 사용하기 때문에 넓고 균일한 영역에는 잡음형태의 움직임 벡터(noisy motion Vector)를 포함하고 있다. 따라서, 본 논문에서는 MPEG2 비디오 스트림에서 완전한 복호화 과정없이 움직임 재예측(motion re-estimation)을 시도하여 움직임 플로우를 구하였으며, 시공간적인 필터를 수행함으로써 근사적인 광 플로우인 움직임 벡터를 재구성하였다.

카메라 움직임의 기술은 어파인(Affine)모델과 원근투영(Perspective-Projection)모델을 사용하는데 2차원 평면의 영상에서 이들 모델의 파라메터를 추정하기 위해서, 일반적으로 단순화된 팬-틸트-줌 모델을 사용한다. 기존의 방법들은 이러한 파라메터를 추출하는 방법으로 최소자승 방법을 사용하였다. 하지만 이러한 방법은 뉴스 비디오의 경우 자막 또는 객체의 움직임에 때문에 전역적인 카메라 움직임을 추정하는데 오류를 내포하게 된다. 즉 재 구성된 광 플로우에는 객체와 카메라의 움직임이 혼재되어 나타난다. 따라서 카메라의 움직임을 분석하기 위해서는 객체의 움직임에 의한 국부적인 광 플로우는 잡음으로 간주되어야 하며 전역적인 카메라 움직임은 이들 객체의 움직임보다 강인해야 한다.

이러한 강인한 추정을 위해서 본 논문에서는 일반화된 Hough 변환 기법을 사용하였다. 일반화된 Hough 변환이란 추정대상을 파라메터화하고 파라메터 공간에서 파라메터의 빈도를 누적시키며, 이 누적된 값을 폴링(polling)하는 최대값 추정기이다. 일반적으로 이러한 최대값 추정기는 평균치를 추정하는 최소자승 방법에 비해 정확한 추정에 장애가 되는 outlier가 존재하는 경우 강인한 추정을 수행할 수 있다[6]. 즉 본 논문의 목적인 영상 전역에 발생하는 카메라 움직임을 추정하는데, 국부적인 객체의 움직임은 outlier로 작용하고, 카메라 움직임에 해당되는 부분에서 최대의 빈도를 갖는다는 현상을 이용하자는 것이다.

이를 위하여 팬과 틸트의 양인(Δx, Δy) 파라메터 공간과 줌의 성도인 α를 나타내기 위해 (중심좌표, 기타) 파라메터 공간을 정의하였고, 파라메터 공간에서 광 플로우의 분포 값으로써 팬-틸트-줌 값을 결정하였다.

제안된 방법은 축구경기 및 농구경기 비디오에 대한 실험결과 93.75%의 성확도를 보였으며, 예상했던 바와 같이 배경이 프레임에 차지하는 부분이 작거나 아주 좁고 빠른 카메라의 움직임이 있어서 전체 프레임에서 움직임 플로우가 극히 작은 부분만 정의되는 경우, 즉 Hough 변환에 의한 추정에 Break Down이 일어나는 경우만을 제외하고는 만족할 만한 성능을 보였다.

제안된 알고리즘은 영화등에서는 감독의 카메라 직입 특징을 분석하거나 카메라 움직임에 공격회수 등의 비중을 부여할 수 있는 농구, 배구 등의 정기 자동분석 등에 이용할 수 있으며, 향후 MPEG2 비디오 스트림에서 복호화 과정없이 카메라의 움직임을 보상한 객체의 움직임 분석에 활용할 수 있을 것으로 기대된다[7].

본 논문의 2절에서는 압축정보를 얻기 위한 MPEG2 비디오 스트림의 구조 및 움직임 벡터의 추출에 관해서 논하였고 3절에서는 근사적인 광 플로우를 위한 움직임 벡터의 재 예측, 잡음 제거를 위한 시공간적인 필터에 대해서 논하였다. 그리고 4절에서는 팬-틸트-줌 카메라 모델과 이를 위한 파라메터 정의 및 추출방법에 대해서 논하였고, 5절에서는 실험영상에 대한 실험 및 실험결과에 대한 카메라의 동작을 고찰하였다. 마지막으로 6절에서는 결론 및 향후계획에 관해서 언급하였다.

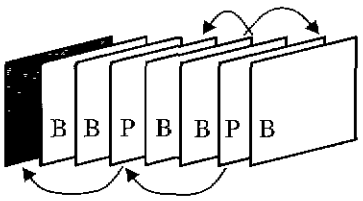
2. MPEG2 비디오 스트림의 분석 및 움직임 벡터의 추출

2.1 스트림의 분석

압축정보의 추출 및 이용방안을 위해 우선 MPEG2

비디오 스트림의 구조를 살펴보고자 한다. MPEG의 압축 기술은 이미지의 시간적 공간적인 유사성 이용하여 중복성을 작게 만들려는 방법으로 압축되어 있는데 카메라 움직임 인식을 위한 압축정보인 움직임 벡터는 시간적인 상관관계를 이용한 방법과 관련이 있으므로 이를 기준으로 설명하고자 한다.

MPEG2에서는 I, P, B 프레임의 종류를 가지고 (그림 1)과 같이 일반적으로 구성되는데, I 프레임은 JPEG과 같아 8×8 블록의 DCT(Discrete Cosine Transform)를 가지고 가변장 부호화(Variable Length Coding)된 이미지로써 GOP(Group of Pictures)의 기준이 되는 이미지이다. P 프레임은 앞의 I 프레임이나 P 프레임으로부터 순방향 예측된 이미지이고 B프레임은 앞뒤의 I 프레임이나 P프레임에서 양방향 예측된 이미지이다. 이러한 프레임 사이의 관계는 GOP 단위 내에서 구성되며 인결된다[8, 9].



(그림 1) 프레임 구조 및 예측 방향

2.2 움직임 벡터의 추출

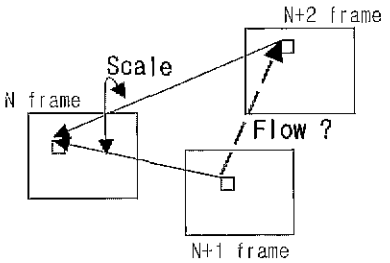
움직임 벡터는 P와 B 프레임에서 매크로 블록단위로 존재하는데 프레임 구조 예측인지 필드구조 예측인지에 따라 2번째 움직임 벡터의 유무가 결정된다[8]. 프레임 구조 예측인 경우는 움직임 벡터의 참조영상이 앞 프레임이므로 이용에 별 문제가 없으나 필드 구조인 경우에는 예측되는 경우에 따라 유용한 움직임 벡터를 선택하여 사용하여야 한다.

즉 (그림 2)는 P프레임에서의 필드구조 예측을 보여주고 있는데 두번째 움직임 벡터를 자신의 프레임에서 예측하는 b)와 c)인 경우에는 자신의 프레임에서의 예측을 하는 두번째 움직임 벡터는 무시하고, 앞 프레임에서 예측하는 첫번째 움직임 벡터만을 취하여 다음 절에서 언급할 광 플로우를 구성하는데 사용한다.

3 움직임 벡터의 활용

3.1 근사적인 광 플로우 구성

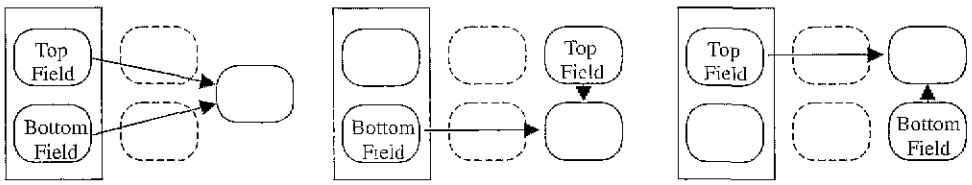
MPEG2에서의 매크로 블록 단위의 움직임 벡터는 압축의 효율을 높이기 위한 예측벡터로 I 프레임에서는 움직임 벡터가 없고, P프레임에서는 바로 전 프레임이 아닌 앞 프레임에서 예측벡터를 취하고, B 프레임에서는 앞뒤에서 양방향예측을 취한다. 이러한 이유 때문에 움직임 측면에서 보면 움직임이 시간적으로 순서적이지 못하고, 예측프레임과의 간격이 달라 움직임 벡터간의 크기척도도 다른 의미를 갖는 문제점이 있다. 즉 (그림 3)과 같이 B 프레임과 B 프레임 사이에서의 시간적인 흐름 움직임 관계를 압축정보에서 직접 얻을 수 없다는 것이다[10, 11].



(그림 3) 광 플로우의 필요성

이러한 이유 때문에 압축 영역에서 직접 얻은 움직임 벡터를 활용하여 프레임별로 순서적인 움직임 벡터를 구하려고 하는데 이를 근사적인 광 플로우라 정의하였다[12].

광 플로우는 각각의 프레임의 종류에 따라 연결할



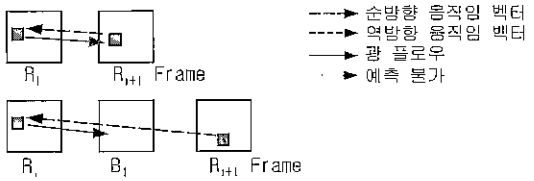
(a) 양쪽가능 (b) Bottom 쪽 가능 (c) Top 쪽 가능

(그림 2) 첫번째와 두번째 필드 예측시 고려할 움직임 벡터

프레임간의 종류가 다음과 같이 4가지 경우가 있을 수 있다.

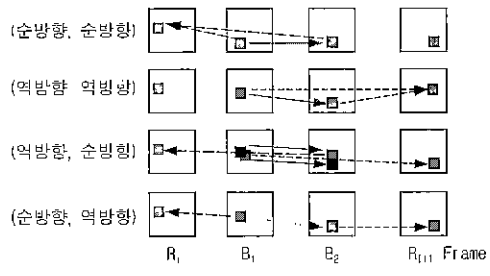
$$\begin{aligned}&\text{(a) } I(t) + P(t+n) \text{ 또는 } P(t) + P(t-n)\\&\text{(b) } B(t) + B(t+1)\\&\text{(c) } B(t) + I(t+1) \text{ 또는 } B(t) + P(t+1)\\&\text{(d) } I(t) + B(t+1) \text{ 또는 } P(t) + B(t+1)\end{aligned} \quad (1)$$

식 (1-a)인 경우 즉 B프레임이 없는 IPPPPP형식으로 구성된 경우로 2번째 프레임의 순방향 움직임 벡터의 반대방향을 첫번째 참조 프레임의 광 플로우로 정의하였다. 즉 (그림 4)는 식 (1-a)의 두 경우에서 광 플로우를 추정하는 방법을 표현하였다. 즉 처음 참조 프레임의 광 플로우는 두번째 참조 프레임의 순방향 움직임 벡터로부터 구하였다. 이때 두 프레임간의 간격이 한 프레임이 아니고, 두 프레임 사이에 B 프레임이 삽입된 경우 프레임의 간격에 따라 움직임 벡터의 크기도 삽입된 프레임수+1 만큼 나누어져야 한다.



(그림 4) 참조 프레임간의 광 플로우 추정

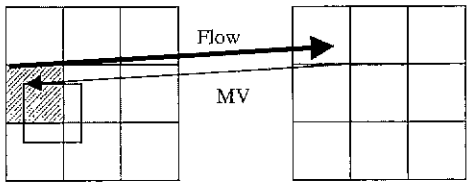
식 (1-b)인 경우 B 프레임 사이에서의 광 플로우를 구하는 방법으로 (그림 5)에 표현하였다. 즉 B 프레임인 경우는 움직임 벡터의 종류에 따라 (순방향, 순방향), (역방향, 순방향), (역방향, 역방향)인 경우에 각 움직임 벡터의 차이값에 의해서 구할 수 있다 물론 (그림 5)의 마지막과 같이 (순방향, 역방향)인 경우에는 광 플로우를 구할수 없다



(그림 5) B 프레임 간의 광 플로우 추정

또한 식 (1-c)인 경우처럼 B프레임과 참조 프레임의 결합은 B프레임의 역방향 움직임 벡터를 광 플로우로 사용하였고, 식 (1-d)인 경우처럼 참조프레임과 B프레임의 결합은 B 프레임의 순방향 움직임 벡터의 반대방향을 광 플로우로 사용하였다.

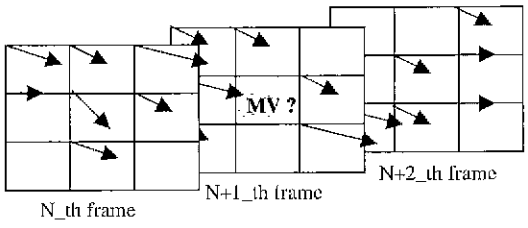
그리고 식 (1)에서 예시한 광 플로우를 생성하는 과정에서 움직임 벡터의 역방향 성분을 갖는 광 플로우를 추정하는 경우, 알맞은 매크로 블록을 찾기 위하여 (그림 6)과 같이 예측된 매크로 블록을 사용하였다. 즉 움직임 벡터의 성분크기 값이 매크로 블록단위와 일치하지 않을 경우 참조된 매크로 블록을 가장 많이 포함하는 빗금 친 매크로 블록이 움직임 벡터의 역 링형의 광 플로우를 갖는 것으로 정의하였다.



(그림 6) 움직임 벡터와 역 방향인 광 플로우의 구성

3.2 시공간적 필터링

MPEG 비디오가 원래의 압축형상을 복적으로 움직임 벡터를 민적상관 방법에 의해 예측하고 부호화 하였으므로, 앞 절에서 구한 광 플로우는 이상적인 광 플로우와는 많은 차이가 있다 이러한 점을 보완하고자 구해진 광 플로우를 평활화할 필요가 있다. 본 논문에서는 (그림 7)과 같이 3*3*3의 크기를 가진 윈도우를 가지고 시간적 공간적인 벡터 메디안 필터링을 취하였다. 벡터 메디안 필터는 벡터 데이터의 불연속을 보존하면서 잡음을 효율적으로 제거하고자, 필터링에 참여하는 벡터들의 중간 값을 갖는 하나의 벡터를 선택하는 비선형 필터이다



(그림 7) 시공간이 3*3*3의 크기를 갖는 윈도우의 벡터 메디안

즉 (그림 7)과 같이 이웃의 움직임 벡터들을 가지고 벡터 메디안 필터를 수행함으로써 잡음형태의 움직임 벡터를 평활화하고 시공간적인 실제의 움직임은 보존하는 역할을 수행하자는 것이다. 이때 갱신된 움직임 벡터값은 식 (2)와 같이 표현될 수 있다.

$$
\begin{aligned}
&if \quad Sum_{ab}^{T} = \sum_{i=-1}^{1}\sum_{j=-1}^{1}\sum_{j=-1}^{1} \mid \vec{M}_{ab}^{T} - \vec{M}_{ij}^{t} \mid \\
&\vec{M}_{nm}^{14} = Median\,Vector\left[Sum_{ab}^{1}, Sum_{cd}^{2}, \cdots, \right. \\
&\qquad \left. Sum_{nm}^{14}, \cdots Sum_{ef}^{25}, Sum_{gh}^{27} \right] \\
&then \;\; \vec{M}_{00}^{0} = \vec{M}_{nm}^{14}
\end{aligned}
\qquad (2)
$$

식 (2)에서 $\vec{M}_{ab}^{T}$는 (그림 7)의 윈도우에서 기준이 되는 (a, b)위치의 움직임 벡터이고 $\vec{M}_{ij}^{t}$는 윈도우안의 다른 움직임 벡터이다 그리고 Sum_{ab}^{T}은 기준 벡터($\vec{M}_{ab}^{T}$)와 윈도우 안의 나머지 벡터들($\vec{M}_{ij}^{t}$)과의 거리의 합을 의미한다. 함수인 MeadianVector()는 각각의 메크로 블록을 기준으로 구한 거리의 합들을 오름차순으로 정렬한 후 중앙값에 해당되는 Sum_{nm}^{14}값의 기준 움직임 벡터인 $\vec{M}_{nm}^{14}$값을 되돌려주는 함수이다. 또한 이 함수에서 Sum값의 윗 첨자는 오름차순의 순위를 의미하고 아래 첨자는 윈도우안의 위치를 의미한다. 그리고 $\vec{M}_{00}^{0}$는 (그림 7)에서 N+1프레임의 갱신될 중심위치의 움직임 벡터를 의미한다.

4. 팬-틸트-줌 모델

4.1 카메라 움직임 모델

카메라 모델은 3차원 공간을 2차원의 이미지로 투영시키는 관계식이다 하지만 역으로 2차원의 이미지에서 3차원의 역투영 변환 모델의 파라메터를 추출한다는 것은 일반적인 MPEG2 비디오 스트림 만으로 구할 수 없다. 따라서 본 논문에서의 카메라 움직임 모델은 일반적으로 사용되는 3차원의 투영모델을 사용하지 않고, 시간에 따라 대응되는 이미지의 좌표 값으로부터 유도되는 근사화된 팬-틸트-줌 모델을 사용하였다. 팬-틸트-줌 모델은 일반적인 원근투영 모델로부터 카메라의 초점거리를 무한하다고 가정하고 카메라의 롤이 없다는 가정 하에서 식 (3)과 같이 표현할 수 있다[13, 14].

$$
\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = S \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \qquad (3)
$$

식 (3)에서 메개 변수 s는 줌 양을 나타내고, t_x, t_y는 팬-틸트량을 의미한다 또한 (x_0, y_0), (x_1, y_1)은 두 영상에서 대응점을 표현한다. 이러한 팬-틸트-줌 파라메터를 계산하기 위해서, 기존의 방법들은 한 프레임의 여러 개의 움직임 벡터를 입력하여 최소한의 오차값을 갖도록 최소자승 알고리즘을 수행함으로써 카메라 모델 파라메터 값을 추정하였다 하지만 이 방법은 추정과정에서 오차를 유발하는 outher에 해당되는 오류블록의 움직임 벡터에 민감하기 때문에 이를 극복하고자 많은 노력을 하였다. 그래서 오류블록을 분류하고 오류벡터를 입력 값에서 제거한 후 최소자승 방법을 이용하는 방법도 제안되었다[7, 10]. 하지만 이 방법 또한 오류블록의 분류가 정확하지 않다는 단점과 오류블록을 제거하는 전처리 단계가 필요하다는 단점이 있다.

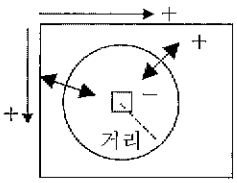
이러한 단점을 극복하고 객체의 국부적인 움직임 또는 뉴스 비디오 등에서 움직이는 자막 때문에 발생하는 잡음에 강인한 카메라 동작을 인식하기 위해서 본 논문에서는 일반화된 Hough 변환 방법을 이용한 파라메터 추정방법을 제안하였다

4.2 일반화된 Hough 변환 방법을 이용한 팬-틸트-줌 양의 계산

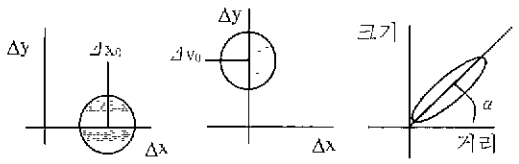
카메라 움직임은 프레임 전역의 광 플로우에 영향을 주며, 객체의 움직임은 객체가 점유하는 부분에만 영향을 준다. 앞서 근사적으로 계산된 광 플로우는 카메라의 움직임 뿐 아니라 국부적인 자막이나 객체의 움직임도 함께 포함되어 있기 때문에 카메라의 움직임은 객체의 움직임에 의해 변화되었다고 볼 수 있다. 따라서 화면 전체에서 일어나는 광 플로우 분포의 최대값은 카메라의 움직임으로 가정할 수 있다

이러한 가정을 토대로 본 논문에서는 팬-틸트-줌 모델의 파라메터 공간을 각 프레임의 광 플로우의 대표적인 Δx_0(팬 양), Δy_0(틸트 양), α(줌 양) 값으로 정의하였다. 이러한 대표 값들은 광 플로우인 움직임 벡터의 분포도를 이용하여 구하였다 (그림 8-a)와 같이 Δx_0, Δy_0는 0값(움직임이 없는 블록)을 제외한 오른쪽과 아래쪽 방향을 양의 축으로 가정하여 움직임 벡터가 가장 많이 발생되는 Δx, Δy 값으로 정의하였다. 그리고 값은 줌의 중심으로 향하는 성분크기를 음의 값으로 밖으로 향하는 성분크기를 양의 값으로 가정하여, 각 움직임 벡터에 대해서 중심과의 거리와 크기 공간에서 분포도를 그린 후 기울기를 구하여 α값으로 정의

하였다 (그림 8)은 Δx_0, Δy_0, α값을 정의하기위한 가정과 이상적인 카메라 동작이 일어날 때 발생되는 Δx_0, Δy_0, α의 움직임 벡터의 분포도를 표현한 그림이다.



(a) Δx_0, Δy_0, α를 정의하기위한 가정

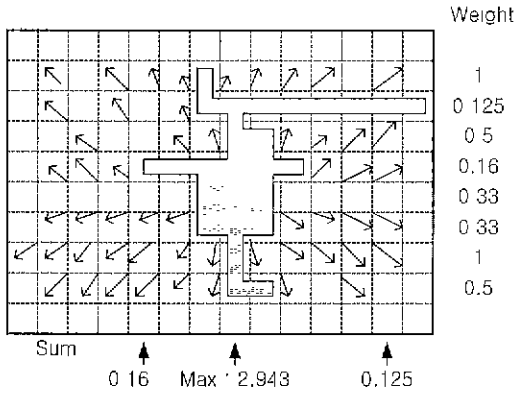


(b) 이상적인 왼쪽 팬, 위쪽 틸트, 줌인 동작

(그림 8) Δx_0, Δy_0, α정의를 위한 가정과 이상적인 카메라 움직임 벡터의 분포도

본 논문의 방법에서 Δx, Δy를 이용한 팬-틸트 양을 구하는 방법은 일반화된 Hough 변환이다. 즉 (Δx, Δy)는 파라메터 공간을 의미하고, 움직임 분포에서 빈도가 가장 많은 부분(최대치)을 팬-틸트 양으로 간주하고 추정하자는 것이다. 또한 (거리, 줌 크기) 공간은 엄밀히 파라메터 공간은 아니지만 기울기(크기/거리)는 1차원 파라메터 공간이라 간주 할 수 있다. 여기서, 파라메터 공간들은 분리된 1차원 어큐뮬레이터들로 구성되기 때문에 일반적으로 Hough 변환이 갖는 메모리 공간의 요구량이 크지 않으며, 매크로 블록 당 Δx, Δy, 거리, 줌 크기 등의 누적과 최대치(mode) 계산에 필요한 계산량 역시 최소자승 방법에 비해 크게 증가하지 않는다

여기서, α값의 정확한 예측을 위해서 줌의 중심 좌표(광학축)를 예측할 필요성이 있다. 왜냐하면 본 논문에서 α값은 줌 중심으로부터의 거리와 광 플로우의 줌 크기로부터 구해지기 때문이다. 이를 위하여 본 논문에서는 양쪽에서 움직임 벡터의 부호가 바뀌고 연속적으로 움직임이 없는 중앙의 매크로 블록들을 대상으로 가중치의 합이 1이 되도록 가로와 세로 방향의 열과 행에 가중치를 주고, 각 행과 열의 가중치를 누적하여 가장 많은 값을 가지는 매크로 블록을 줌의 중심좌표로 정하였다



(그림 9) 줌 중심을 찾기 위한 가중치 마스크

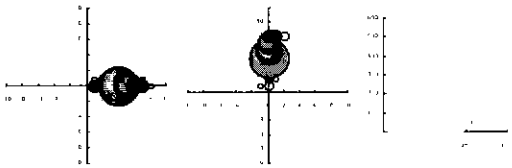
예를 들어 (그림 9)는 줌인(Zoom In)이 발생할 때의 움직임 벡터라고 하면 둘째 라인은 빈 매크로 블록이 1개이므로 가중치1을, 셋째 라인은 8개 이므로 매크로 블록 당 0.125(1/8)의 가중치를 주었다. 이와 같은 방법으로 빈 매크로 블록에 가중치를 준 후에 세로 방향으로 누적하면 8번째 열에서 가장 큰 값인 2.943을 구할 수 있다. 즉, 8열이 줌 중심의 x좌표 값이 되는 것이다 y좌표도 이와 같은 방법으로 구할 수 있게 된다.

5. 실험 및 고찰

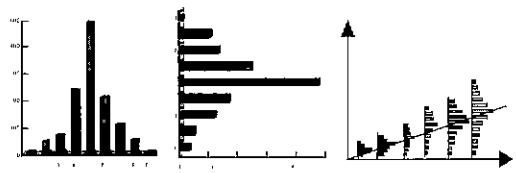
본 논문에서는 순서적인 흐름으로 만들기 위해 압축 영역에서 바로 얻을 수 있는 움직임 벡터를 근사적인 광 플로우로 재구성하였고, 이러한 움직임 벡터의 분포를 이용하여 팬-틸트-줌의 카메라 모델을 정의하였다.

실험에서 움직임 벡터는 이상적인 분포를 가지지 않는데 이는 압축과정에서의 단순히 민적상관 방법에 의해서 예측하고 객체의 움직임을 포함하고 있기 때문이다. 그래서 팬-틸트-줌이 발생하는 실제 영상의 (Δx, Δy) 파라메터 공간과 (거리, 줌 크기) 파라메터 공간에서의 움직임 벡터 분포는 (그림 10-a)와 같이 나타난다 이러한 분포값을 추정하기 위해서 (그림 10-b)와 같이 Δx_0, Δy_0값은 Δx축과 Δy축의 누적된 분산의 최대값으로 추정하였고 α값은 이상적인 직선형태가 아닌 삼각형 형태로 나오지만 분포의 특성은 직선의 형태이므로 기울기인 α값도 같은 방법으로 구할 수 있었다. 이는 실험에서는 팬-틸트-줌을 찾기위해 1차원 어큐뮬

레이더 3개를 사용한 것에 해당되고, 메모리 요구량과 최대값 추정의 계산량은 더욱 줄어들 수 있음을 의미한다.



(a) 실제영상에서 왼쪽 팬, 윗쪽 틸트, 줌인이 발생할 때 움직임 벡터 분포

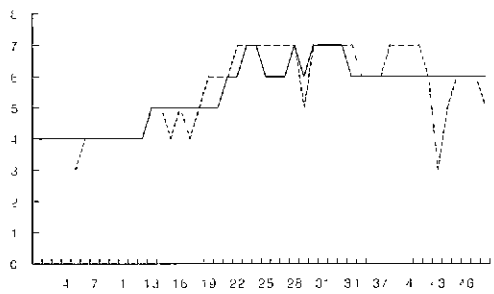


(b) Δx_0, Δy_0, α를 구하기 위한 히스토그램

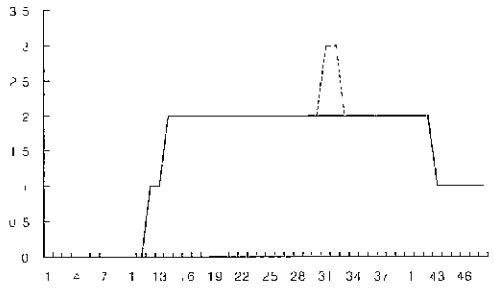
(그림 10) 실제 영상의 움직임 벡터 분포도 및 히스토그램을 이용해 구한 Δx_0, Δy_0, α

제 구성된 광 플로우로부터 구해진 팬-틸트-줌 카메라 모델의 파라메터를 비교하기 위해서, 기존의 최소자승 방법과 본 논문에서 제안한 Hough 변환 방법을 사용하여 추출된 결과를 비교하여 보았다. 앞절에서 언급한 것처럼 카메라의 움직임만 프레임에 전역적으로 분포하는 경우는 최소자승 방법과 제안된 Hough 변환 방법은 (그림 11)과 같이 유사한 결과를 보였다. (그림 11)의 가로축은 프레임 수를 의미하며 세로축은 각 파라메터((a) Δx_0, (b) Δy_0, (c) α)의 크기를 의미한다.

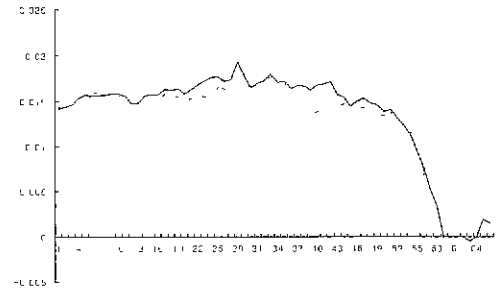
하지만 (그림 12)와 같이 자막이 있거나 국부적인 움직임 객체가 있는 부분에서는 제안된 Hough 변환방법이 오류에 강인하게 카메라 동작을 추정함을 알 수 있었다. 즉 (그림 12-b)에서 x는 전체 프레임에 대한 기존의 최소자승 방법에 의한 카메라 움직임 추정 결과이고, y는 오류인 벡터가 있는 자막을 제거한 프레임에 대한 최소자승 방법에 의한 카메라 움직임 추정 결과이다. 그리고 z는 전체 프레임에 대한 제안된 일반화된 Hough변환 방법에 의한 카메라 움직임 추정이다. 결과적으로 기존의 최소자승 방법보다 제안된 Hough 방법이 오류가 없는 결과값과 유사한 결과를 추정함을 알 수 있었다.



(a) 왼쪽 팬 동작

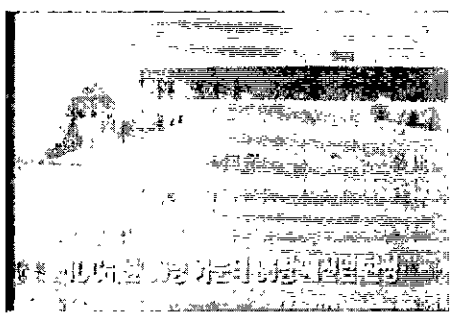


(b) 윗쪽 틸트 동작

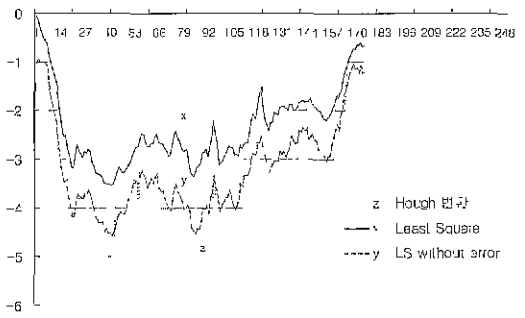


(c) 줌인 동작

(그림 11) 최소 자승 방법과 제안된 Hough변환 방법의 비교



(a) 자막이 있는 오른쪽 팬 동작의 영상



(b) 카메라 동작 파라메터의 비교

(그림 12) 오류벡터가 있는 영상에서의 최소자승 방법과 제안된 Hough변환 방법의 비교

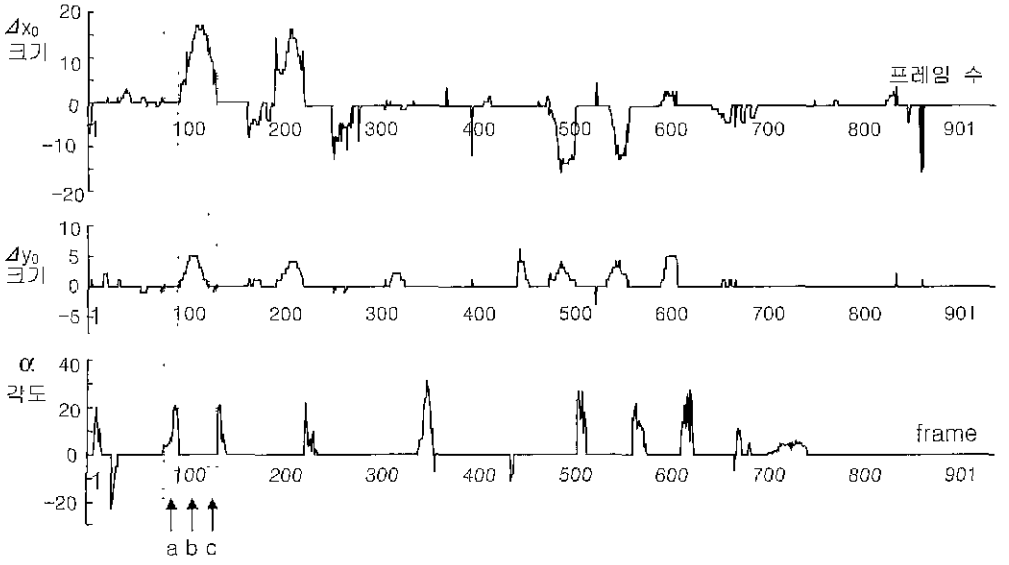
제안된 방법에 의해 추정된 파라메터를 통해 카메라의 동작을 분석하기 위해서, (그림 13)은 MPEG2로 압축된 704*480크기의 약 1000 프레임의 농구경기 실험영상에서 각 프레임마다 추출된 Δx_0, Δy_0, α를 가지고 카메라의 동작을 추정한 결과를 도시하였다. 결과에서 보듯이 단일 카메라의 움직임 뿐만 아니라 다중의 카메라의 움직임도 해석할 수 있었다. 즉 (그림 13)의 일부 구간인 a구간은 α값이 양의 값이므로 줌인 동작이고, b 구간의 Δx_0의 Δy_0값들이 양의 값이므로 카메라가 윗쪽과 왼쪽으로 움직이는 팬-틸트 동작이다. 그리고 c구간에서는 Δx_0값들이 양의 값이므로 카메라는 왼쪽으로 움직이는 팬 동작임을 알 수 있다.

실험영상의 프레임 중 (그림 13)의 abc구간의 프레임을 복호화하여 살펴보면 추정한 결과와 같음을 (그림 14)를 통해서 알 수 있다. 그리고 실험영상의 전체 프레임에 대한 분석 결과를 <표 1>에서 비교하였다.

실험의 정확성을 위해서 다른 실험 비디오인 농구경기와 축구경기의 비디오에 대해서도 같은 방법으로 실험한 결과, <표 2>에서와 같은 결과를 얻을 수 있었다.

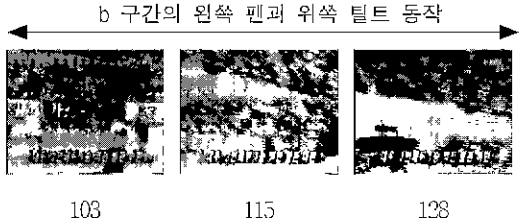
<표 1>과 <표 2>에서 참값은 사람이 판별한 카메라 움직임이 발생한 프레임의 수이고, 측정값은 실험에 의해서 추정된 결과값이다. 그리고 Miss는 카메라의 움직임을 인식하지 못한 경우이고 False Alarm은 올바른 카메라 움직임이 아닌 것을 카메라 동작으로 잘못 인식한 경우를 의미한다. 그리고 정확률은 전체 프레임에서 바르게 찾은 프레임의 비율인 (총프레임수 - False Alarm-Miss)/총프레임수*100(%) 값으로 계산하였다.

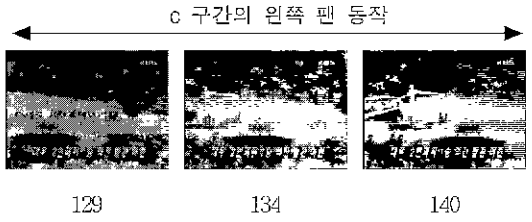
실험 결과에서 오인식은 예상했던 바와 같이 카메라가 너무 빨리 움직여서 배경이 평활화되어 배경의 움직임 벡터가 소멸되거나, 움직이는 객체가 너무 커서 배경의 움직임이 극히 작게 발생하는 경우 발생하였다. 즉 Hough 변환의 Break Down이 일어나는 경우에 해당된다고 볼 수 있다. <표 2>의 실험에서 이러한 오류들은 극히 작은 값으로 나타났고, 올바르게 인식한

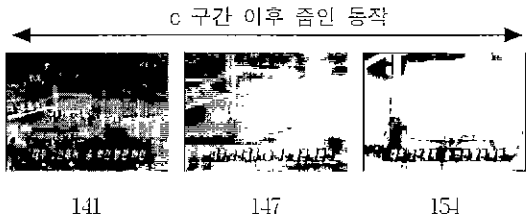


(그림 13) 실험영상에서의 프레임의 대표값인 Δx_0, Δy_0, α 의 변화량

프레임의 인식률은 전체 총 11,529프레임 중에서 93.75%인 10,808프레임을 올바르게 인식하였다







(그림 14) 카메라 움직임을 획인하기 위한 복호화된 실험 비디오의 일부구간

〈표 1〉 농구 실험 비디오에 대한 각각의 카메라 동작에 대한 인식률

	오른쪽 팬	왼쪽 팬	윗쪽 틸트	아래쪽 틸트	줌 인	줌 아웃	합계
참 값	148	133	57	13	156	17	524
측정값	130	119	51	11	141	14	469
Miss	22	18	9	2	15	3	69
False Alarm	4	1	3	0	3	0	11
정확률(%)	85.1	86.5	81.2	84.6	90.3	92.4	91.16
총프레임	939						

〈표 2〉 여러 실험 비디오에 대한 카메라 움직임 인식결과

	총프레임 수	참값	측정값	Miss	False Alarm	정확률(%)
농구1	1121	618	575	61	18	92.95
농구2	1208	318	289	36	7	96.44
농구3	835	485	447	54	16	92.09
농구4	1691	995	948	73	26	94.15
축구1	1829	938	856	103	21	93.22
축구2	946	250	217	43	10	91.40
축구3	2379	762	682	106	26	91.45
축구4	1470	493	429	92	29	91.70

6. 결 론

본 논문에서는 MPEG의 압축영역에서 얻을 수 있는 움직임 벡터를 가지고 흐름이 순차적인 광 플로우로 갱신한 후, 일반화된 Hough 변환을 이용하여 카메라 움직임 모델의 파리메터를 추정하였다. 본 논문에서 적용한 Hough 변환방법은 카메라의 움직임이 프레임의 전부분에 발생하고, 예측의 잡음이나 자막 또는 객체의 움직임은 국부적으로 발생하기 때문에 카메라의 움직임이 이러한 잡음 벡터에 의해 변화된다는 점에 기초하였다. 본 논문에서 가정한 카메라 움직임은 팬-틸트-줌 모델이며, 이들 모델의 매개변수는 Δx_0, Δy_0, α에 의해 표현된다. 제안된 Hough변환 방법은 기존의 최소자승 방법에 비해 오류에 강인하였다. 이러한 이유 때문에 오류를 극복하고자 오류인 블록을 제거하는 전처리 과정을 생략할 수 있었다. 또한 압축영역의 정보를 이용함에도 불구하고 농구경기, 축구경기 등의 MPEG2 비디오 스트림에 적용해 본 결과, 카메라 동작이 있는 부분에서만 87.5% 정확성과 카메라 동작이 없는 부분에서 거의 100%의 정확성을 가진 측정치를 보였다. 이는 전체적으로 93.75%이상의 높은 정확성을 보인다.

제안된 알고리즘은 영화등에서는 감독의 카메라 작업 특징을 분석하거나 카메라 움직임에 의미를 부여할 수 있는 농구, 배구 등의 경기 자동분석등에 이용할 수 있으며, 향후 MPEG2 비디오 스트림에서 복호화 과정없이 카메라의 움직임을 보상한 객체의 움직임 분석에 활용할 수 있을 것으로 기대된다. 또한 압축 비디오를 검색하기 위한 속성인지로써 기존에 사용되던 매크로 블록 정보나 색상정보, 위치정보, DCT의 질감 성분등과 더불어 중요한 속성의 하나로 활용될 것이라

기대된다.

향후 연구계획은 추출된 카메라의 움직임을 바탕으로 움직임이 보상된 객체를 인식하고 추적하거나, 카메라 파라메터를 이용하여 대표 프레임으로 이용되는 모자이크 영상을 구현할 것이다.

참 고 문 헌

[1] V Kobla, and D Doermann, "Compressed domain video indexing techniques using DCT and motion vector information in MPEG video," *Storage and Retrieval for Image and Video Databases V*, SPIE Vol.3022, pp.200-211, Jan 1997.

[2] O. N. Gerek, Y. Altunbasak, "Key Frame Selection from MPEG Video Data," *VCIP*, SPIE Vol.3024, pp.920-925, 1997

[3] J Meng, Y. Juan, S-F Chang, "Scene Change Detection in a MPEG Compressed Video Sequence," Digital Video Compression; Algorithms and Technologies, SPIE Vol 2419, pp.14-25, 1995.

[4] Y Ariki, Y Saito, "Extraction of TV News Articles Based on Scene Cut Detection Using DCT Clustering," *ICIP 96*, Vol 3, pp.847-850, 1996.

[5] Marco la Cascia, Edoardo Ardizzone, "JACOB : Just a Content-Based Query System for Video Databases," *ICASSP'96*, Vol 2, pp 1216-1219, May 1996.

[6] Ramesh Jain, Rangachar Kasturi and Brian G Schunck, 'Machine Vision', McGraw-Hill, 1995.

[7] Saur Drew D, Tan Yap-Peng, Kulkarni Sanjeev R, Ramadge Peter J, "Automated Analysis and Annotation of Basketball Video," *Storage and Retrieval for Image and Video Databases V*, SPIE Vol. 3022, pp.176-187, Jan 1997

[8] ISO-IEC13812-1/ISO-IEC13812-2 International Standards, 1st Ed, 1996.

[9] Keith Jack, 'Video Demystified A Handbook for the Digital Engineer', 2nd Ed., Hightext Interactive, 1996.

[10] R. Milanese, F. Deguillaume, A Jacot-Descombes, "Efficient Segmentation and Camera Motion Indexing of Compressed Video," *Real-Time Imaging*, Vol.5, No.4, pp.231-241, Aug 1999.

[11] Maurizio Pilu, "On using raw MPEG motion vectors to determine global camera motion," *Visual Communications and Image Processing '98*, SPIE Vol.3309 pp 448-459, Jan 1998.

[12] P. Sobey, M. V. Srinvansan, "Measurement of Optical Flow by Generalized Gradient Scheme," *J.Opt Soc.Am.A*, Vol.8, No.9, pp.1488-1498, 1991.

[13] Isaac Cohen, Gerard Medioni, "Detection and Tracking of Objects in Airborne Video Imagery," *CVPR'98 Workshop on Interpretation of Visual Motion*, 1998. http://iris.usc.edu/home/raycharles/_Mosaic/Outlines/Papers.html

[14] A Murat Tekalp, 'Digital Video Processing', Prentice Hall, 1995

유 원 영

e-mail : zero2@iceng.chonbuk.ac.kr

1996년 전북대학교 전자공학과 졸업(학사)

1998년 전북대학교 대학원 영상정보공학과 졸업(공학석사)

1998년~현재 전북대학교 대학원 전자공학과 박사과정

관심분야 비디오 분석 및 편집, 영상처리

최 정 일

e-mail : jichoi@ailab.chonbuk.ac.kr

1999년 전북대학교 전자공학과 졸업(학사)

1999년~현재 전북대학교 대학원 전자공학과 석사과정

관심분야 비디오 분석 및 영상처리

이 준 환

e-mail : chlee@moak.chonbuk.ac.kr

1980년 한양대학교 전자공학과 졸업(학사)

1982년 한국과학기술원 전자공학과 졸업(공학석사)

1990년 미주리대 전기 및 컴퓨터 공학과 졸업(공학박사)

1999년 현재 전북대학교 전자공학과 정교수

관심분야 : 영상처리 및 분석, 인공지능