

MPEG 압축 정보를 이용한 모자이크 구성

정희원 설 정 규*, 이 승 희**, 이 준 환***

A Composition of Mosaic Images based on MPEG Compressed Information

Jung-kyu Seol*, Seung-hee Lee**, Joonwhoan Lee*** *Regular Members*

요 약

본 논문에서는 압축된 MPEG2 비디오 스트림이 포함하고 있는 정보들을 이용하여 카메라의 움직임에 따른 연속적인 프레임의 변위를 추정하고 이를 기반으로 모자이크 이미지를 구성하는 방법을 제안한다. 제안된 방법에서는 MPEG의 매크로 블록 단위의 움직임 벡터로부터 근사적인 광 플로우(Optical Flow)를 구성하고, 이를 이용하여 팬(Pan), 틸트(Tilt) 등의 카메라 동작에 따른 프레임의 변위량을 추출한다. 추정된 변위량은 모자이크 이미지를 구성하는데 있어서 연속되는 비디오 프레임의 기하학적인 변환을 결정한다.

본 논문의 모자이크 구성은 Nicolas가 제안한 분석적 가중치 등을 이용한 혼합(blending) 방법을 사용하였다. 실험결과 분석적인 가중치를 이용한 혼합방법은 평균이나 미디안에 의한 혼합방법보다 배경의 급격한 변화가 적으며 비디오 프레임의 정보를 충실히 활용함을 알수 있었다. 구성된 모자이크는 압축정보를 디코딩하고 정확한 광 플로우를 구성하는 방식에 비해 부호화 과정에 포함된 움직임 벡터를 이용하여 매크로 블록 단위의 근사적인 광 플로우를 이용하기 때문에 계산량을 줄일 수 있다. 구성된 근사적인 모자이크는 비디오의 정보를 요약하고 있기 때문에 압축된 비디오 검색을 위한 대표 프레임 등에 사용될 수 있다.

ABSTRACT

This paper proposes a composition method of mosaic image from the compressed MPEG-2 video stream, in which the displacement between successive frames according to the camera operation is estimated directly from the information involved in the stream. In the proposed method, the approximated optical flow is constructed from motion vectors of macro blocks, and it is used to determine the parameters of the displacements according to the camera operation associated with pan and tilt. The extracted parameters are used to determine the *geometric transform of successive video frames in order to construct a mosaic image*.

The construction of mosaic uses several blending techniques including the one proposed by Nicolas in which an analytic weight is used to determine pixel values. Through the experiment, the blending technique based on analytic weights was superior to the others such as averaging and median-based techniques. It provided more smooth changes in background and made use of instantaneous frame information to construct a mosaic. The mosaic in the paper puts the emphasis on the reduction of computation because it is constructed from the motion vectors included in the compressed video without decoding and recalculating exact optical flows. The constructed mosaic can be used in the retrieval of the compressed video as the representative frame of a shot.

* 전북대학교 전자공학과 인공지능 연구실 (semlover@moak.chonbuk.ac.kr),

** (dream844@hotmail.com),

*** (chlee@moak.chonbuk.ac.kr)

논문번호 020440-1011, 접수일자 2002년 10월 11일

I. 서론

컴퓨터의 발달과 인터넷의 확산은 기존의 텍스트 위주의 콘텐츠를 영화, 스포츠 중계, 강의 등의 동영상 포함 멀티미디어 콘텐츠로 급속히 변화시키고 있다. 이러한 멀티미디어 데이터는 대부분 데이터의 양이 방대하기 때문에 저장 및 전송효율을 위하여 압축되어지며, 특별히 동영상 데이터는 MPEG-1, MPEG-2등의 방식의 표준화된 압축방식을 이용하고 있다.

본 논문에서는 MPEG 형식으로 되어있는 동영상 데이터의 일련의 프레임으로 이루어진 샷(shot)로부터 모자이크 영상을 재구성하는 문제를 다룬다. 일반적으로 동영상으로부터 구성된 모자이크는 동영상 압축, 편집 등에 이용될 수 있으며, 경우에 따라서는 동영상의 한 샷을 대표하는 대표 프레임으로 활용될 수도 있다^{[1][2]}. 본 논문에서는 후자의 목적을 위해 모자이크를 구성하기 때문에 모자이크의 압축 및 복원, 모자이크 상에서의 비디오 편집 및 복원 등에 수반되는 모자이크 구성의 정확성 측면보다는 대략적인 동영상 데이터의 개요를 빠른 시간에 파악하는데 초점을 둔다^{[3][5]}. 이러한 이유 때문에 본 논문의 모자이크 구성은 정확한 광 플로우의 계산보다는 MPEG 데이터가 포함하고 있는 움직임 벡터를 이용하여 근사화된 광 플로우를 계산하고, 이를 이용하여 카메라의 움직임에 해당하는 프레임 간의 전역적인 변위량을 계산한다. 동영상의 한 프레임에서 움직이는 객체가 차지하는 부분이 상대적으로 작은 부분이라면 Hough 변환기법은 안정적인 프레임간의 기하학적인 변환을 제공해 준다^[6]. 본 논문에서 이러한 안정적인 예측기를 통하여 얻어진 카메라 움직임에 따른 변위량들은 모자이크 구성의 워핑(Warping)에 이용되고, 디코딩된 프레임은 Nicolas가 제안한 프레임 지연과 프레임 사이의 불연속을 최소화하는 방법 등에 의해 혼합하였다^[7].

이러한 방법으로 구성된 모자이크는 압축 비디오가 가지고 있는 정보를 직접 사용하기 때문에 광 플로우를 구하기 위해 소요되는 계산시간을 절약할 수 있으며, Hough 변환에 의해 안정적인 프레임 간의 전역적인 움직임을 예측할 수 있고, 또한 구성된 모자이크는 비디오 샷의 개요를 파악하는데 충분하게 이용될 수 있다. 따라서 본 논문의 모자이크는 압축된 비디오에서 원하는 장면의 검색을 위한 대표 프레임을 빠른 시간에 구성하는데 활용될 수 있을

것으로 기대된다.

본 논문의 2 절에서는 모자이크 구성을 위한 카메라 움직임에 기인한 변위 파라미터 추출과 모자이크 구성방법에 대하여 기술하였으며, 3 절에서는 실험을 통해 대표 프레임으로 사용가능함을 보였다. 마지막으로 4 절에서는 본 논문의 결론 및 향후 계획에 대해서 논하였다.

II. 모자이크 구성

모자이크는 그 구성방법에 따라 특정 프레임을 뽑아 각각의 프레임들 간의 카메라 움직임 정보를 이용하여 만들어진 정적인 모자이크(static mosaic)와 비디오의 연속된 프레임을 이용하여 점진적으로 모자이크를 완성해가는 다이내믹 모자이크 등이 있다^{[2][7][8]}. 본 논문의 모자이크 구성은 압축된 동영상으로부터 그림 1과 같은 절차에 의해 구성되는 다이내믹 모자이크이다.

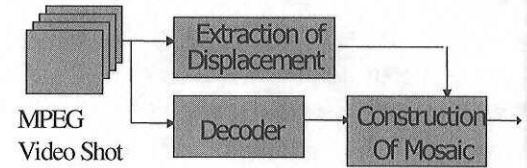


그림 1. 모자이크 구성 절차

그림 1에서 제안된 방식의 입력은 압축영역에서 샷 단위로 구분된 MPEG 압축 비디오라고 가정한다^[4]. 카메라 움직임에 기인한 광 플로우의 계산은 MPEG 비디오 데이터의 움직임 벡터를 기반으로 매크로 블록 단위로 이루어지며 움직임 벡터가 정의되지 않은 블록의 보간과 사공간적으로 불규칙한 움직임 벡터들을 평활화(smoothing)하기 위하여 베디안 필터를 통과시킨다. 필터링 된 광 플로우에서 카메라의 움직임으로부터 기인한 프레임 사이의 변위를 추정하기 위해서는 Hough 변환을 이용한다^[6]. 이와 같은 방법으로 추정된 변위는 매 프레임 단위로 예측되며, 변위량은 디코딩 된 프레임들을 연속적으로 접합하여 모자이크를 구성하는데 이용된다.

2.1 모자이크 구성을 위한 변위추출

제안된 방법에서는 MPEG 비디오 스트림의 매크로 블록단위의 움직임 벡터를 이용하여 복호화 과정 없이 근사적인 광플로우를 구성하고, 이를 이용하여 프레임 사이에 변위량을 추출한다.

가. 근사적인 광 플로우의 구성⁽⁶⁾

MPEG 비디오의 움직임 벡터는 압축효율을 높이기 위해 사용되기 때문에 이들 성분들이 순서적이지 못하고, 공간적으로 균질(homogenous)한 영역에서는 움직임 벡터들이 불규칙하여 카메라 또는 객체의 움직임을 잘 반영하지 못한다. 또한 모든 매크로 블록이 움직임 벡터를 가지는 것은 아니다. 때문에 움직임 벡터들로부터 근사적인 광 플로우를 계산하기 위해서는 먼저 각 프레임의 압축 방법에 따라 또는 움직임 벡터의 성질에 따라 순서적으로 재배열하고, 배열된 벡터들을 필터링하여 난잡한 움직임을 평활화 하고 움직임이 정의되지 않은 블록이 가능한 적어지도록 하는 보간 과정이 필요하다.

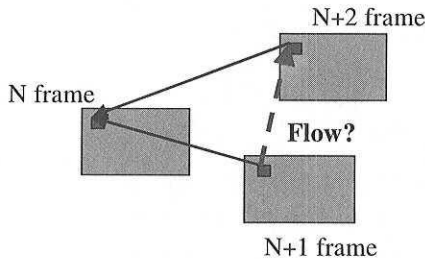


그림 2. 매크로 블록단위의 광 플로우 예측

그림 2는 매크로 블록단위의 움직임 벡터로부터 광 플로우 예측의 필요성을 보여주고 있다. 그림에서 N frame은 참조 프레임(reference frame)을 표현하며 N+1, N+2 프레임이 참조 프레임을 참조하여 실선과 같은 움직임 벡터를 가진다고 가정할 경우 참조 프레임에서 N+1 프레임으로의 광 플로우는 움직임 벡터(실선)의 반대방향으로 움직임 벡터와 같은 크기로 가정할 수 있다. 그러나, N+1 프레임에서 N+2 프레임으로의 광 플로우는 움직임 벡터로부터 직접 얻어낼 수 없고 두 움직임 벡터의 차를 이용할 수밖에 없다.

광 플로우의 예측을 위해 프레임의 연결에 따라 고려할 형태는 다음의 4 종류가 있다⁽⁹⁾.

- (a) $I(t) + P(t+n)$ 또는 $P(t) + P(t+n)$
- (b) $B(t) + B(t+1)$
- (c) $B(t) + I(t+1)$ 또는 $B(t) + P(t+1)$ (1)
- (d) $I(t) + B(t+1)$ 또는 $P(t) + B(t+1)$

식 (1-a)에서의 경우는 그림 3에 표현된 바와 같다. 그림 3의 첫 번째 경우는 순방향 예측된 움직임

벡터의 반대방향을 광 플로우로 가정할 수 있는 경우를 의미하고, 두 번째 경우는 두 참조 프레임 사이에 B 프레임이 삽입된 경우로 움직임 벡터의 크기를 삽입된 프레임수+1로 나누고 방향을 반대로 표현하는 경우이다.

그림 4는 (1-b)의 경우로 움직임 벡터의 종류에 따라 (순방향, 순방향), (역방향, 순방향), (역방향, 역방향)인 경우에 각 움직임 벡터의 차이에 의해 광 플로우를 구성할 수 있다. 그러나, 그림 4의 마지막에 표현된 (순방향, 역방향)인 경우에는 광 플로우를 구성할 수 없다.

식 (1-c)의 경우에는 B 프레임이 참조한 프레임의 움직임 벡터는 방향 전환 없이 그대로 광 플로우로 가정하였으며, 식 (1-d)의 경우에는 B 프레임의 순방향 움직임 벡터의 반대방향을 움직임 벡터로 가정하였다.

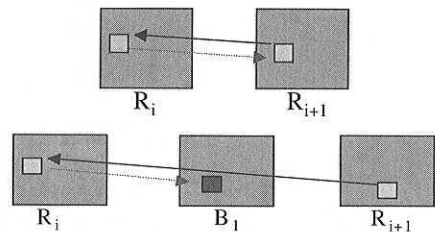


그림 3. 참조 프레임간의 광 플로우 추정

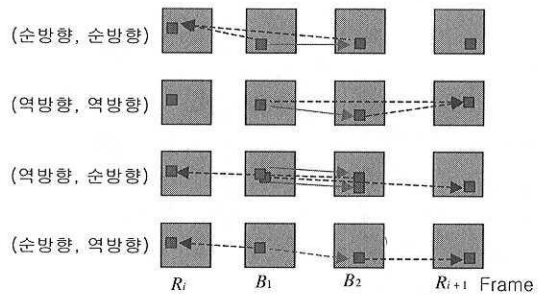


그림 4. B 프레임간의 광 플로우 추정

또한 식 (1)에서 예시한 광 플로우를 생성하는 과정에서 움직임 벡터의 역방향 성분에 따라 해당 매크로 블록을 이동하면 참조 프레임의 매크로 블록과 일반적으로 일치하지 않는다. 이러한 경우에는 참조 프레임에 가장 많이 겹치는 그림 5의 매크로 블록이 해당 움직임 벡터의 역방향의 광 플로우를 갖는다고 가정하였다.

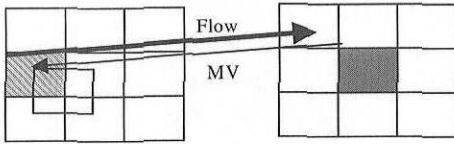


그림 5. 움직임 벡터와 역 방향인 광 플로우의 구성

나. 광 플로우의 보간 및 평활화

전술한 방법으로 구성된 매크로 블록단위의 광 플로우는 MPEG 비디오가 압축향상을 목적으로 움직임 벡터를 예측하고 부호화 하기 때문에 시공간적으로 많은 잡음을 포함하고 있으며, 광 플로우가 정의되지 않는 매크로 블록을 포함할 수 있다.

이러한 시공간적인 잡음을 제거하고, 광 플로우가 정의되지 않는 블록을 보간하기 위해 그림 6과 같이 3*3 크기의 윈도우를 이용하여 시공간적인 메디안 필터링을 수행하였다. 이러한 메디안 필터의 결과는 시공간적인 움직임의 불연속을 보존하면서 불규칙한 잡음 형태의 광 플로우를 평활화 할 수 있다.

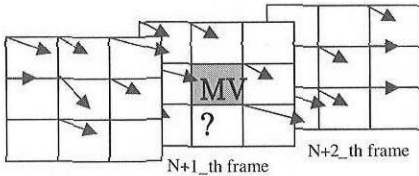


그림 6. 시공간적인 메디안 필터링

그림 6의 필터링 대상은 광 플로우에 해당하는 벡터정보로서 메디안 벡터는 해당 윈도우 내의 광 플로우 벡터간의 유클리디안 거리의 합을 최소화 하는 벡터에 해당한다. 이러한 총체화 정렬(aggregate ordering)방식을 이용한 벡터정보의 순서통계(order statistics) 필터링은 벡터의 불연속을 보존하며 가우시안 형태의 잡음을 효과적으로 제거할 수 있다.

다. 연속적인 워핑(warping)을 위한 변위 추출

동영상으로부터 모자이크를 구성하기 위해서는 연속되는 프레임 사이의 워핑을 위한 파라미터를 추출하여야 한다. 이러한 파라미터를 추출하는 방법에는 카메라의 움직임을 가정하는 방식과 영상의 변위 및 줌에 해당하는 파라미터를 직접 추출하는 방식이 있다. 전자의 경우에는 예측된 카메라의 움직임으로부터 모자이크 구성에 기여하는 영상의 변

위를 구하며, 후자의 경우에는 2차원 영상사이의 변환과 해당 파라미터를 직접 얻어서 워핑에 활용한다. 제한한 방식의 경우에 있어서는 카메라의 움직임을 알 수 있는 정보가 제한된 영상사이의 매크로 블록 단위의 근사화 된 움직임이기 때문에 후자의 방법을 택하였다.

줌과 롤 및 축 방향의 변위를 가정한 일반적인 2차원 영상의 변환은

$$\vec{d} = \begin{pmatrix} \Delta x + a(x-x_0) - \theta(y-y_0) \\ \Delta y + a(y-y_0) - \theta(x-x_0) \end{pmatrix} \quad (2)$$

와 같이 표현된다. 여기서 장면의 배경이 카메라로부터 충분히 멀리 떨어져 있다면, 작은 카메라의 펜과 킬트는 2차원 영상의 축 방향의 변위로 취급할 수 있다. 또한 카메라의 롤을 무시하면 파라미터인 θ 를 제외할 수 있으며, 식 (2)는

$$\vec{d} = \begin{pmatrix} \Delta x + a(x-x_0) \\ \Delta y + a(y-y_0) \end{pmatrix} \quad (3)$$

과 같다. 본 논문에서는 근사적인 광 플로우를 이용하기 때문에 줌 동작에 따른 모자이크 구성은 어려움이 있으며, 따라서 연속되는 영상의 변위에 의한 모자이크만을 고려하여, 식 (3)에서 $a=0$ 인 경우를 가정하였다. 모자이크 구성에 기여하는 연속된 두 장의 프레임으로부터 식 (3)의 파라미터 $\Lambda = (\Delta x, \Delta y, a)$ 를 구하기 위해서는

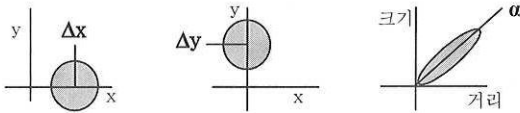
$$I_{t+1}(\vec{b}) \approx I_t(\vec{b} + \vec{d}(\Lambda)) \quad (4)$$

를 최대로 만족하는 Λ 를 찾아야 하며, 이를 위하여 최소자승 방법(least square method)을 사용할 수 있다. 그러나 최소자승 방법에서는 한 프레임 내에서 일관성 없는 광 플로우가 outlier로 작용하여 파라미터 추정에 악영향을 줄 수 있다. 따라서 outlier에 해당하는 움직임은 객체 등에 기인한 광 플로우 부분은 제거하고 배경에 해당하는 부분만 추출하여 파라미터 추정에 이용할 수 있는데 이 경우 움직임은 객체에 대한 선지식을 요구하는 전처리 단계가 필요하다¹⁰⁾.

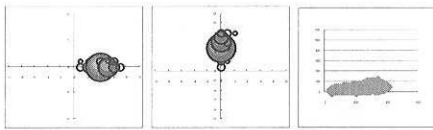
outlier에 따라 추정오차가 발생하는 문제점과 선지식이 요구되는 전처리 단계를 포함하는 단점을 보완하는 방법으로 본 논문에서는 가장 강건한(robust) 파라미터 추정방식인 Hough 변환을 이용한다. 일반적으로 Hough 변환은 계산량이 많으나 고려한 모델에서는 줌과 변위가 동시에 일어나지

않는다고 가정할 경우 이들 파라메터들이 결합되어 있지 않기 때문에 일차원 파라메터를 가정하고, 파라메타 값의 빈도를 이용하여 쉽게 그 값을 추정할 수 있다.

그림 7(a)는 왼쪽으로의 변위(편의상 팬)과 위쪽으로의 변위(편의상 틸트)량의 분포를 1차원 어큐멀레이터 공간에 표현한 것으로 3차원 공간에 구성되지 않기 때문에 계산 량과 메모리 요구 량은 많지 않다.



(a) 이상적인 매개변수 분포도



(b) 이상적인 매개변수 분포도

그림 7. 팬 틸트 줌의 매개변수 분포도

그러나, Hough 변환시의 Δx , Δy 의 해상도는 1 픽셀 단위이기 때문에 경우에 따라 문제를 발생시킬 수 있다. 본 논문에서는 연속되는 프레임으로부터 이들 파라메터를 구하기 때문에 연속되는 프레임의 변위가 1 픽셀 보다 작은 경우가 계속되면 실제 변위가 있음에도 불구하고 Hough 변환에 의해 구해지는 변위는 모두 0이다. 즉 누적되는 작은 변위량이 잘라버림에 의해 모자이크 구성에 반영되지 못하는 경우가 발생한다. 이러한 문제를 해결하기 위해서 본 논문에서는 어큐멀레이터에서 최고 치에 해당하는 값뿐만 아니라 상위 3개의 최고 치에 해당하는 변위량들을 취하고 (5)같이 이들의 가중

$$\Delta = \frac{(w_1 b_1 + w_2 b_2 + w_3 b_3)}{w_1 + w_2 + w_3} \quad (5)$$

평균으로 실수범위의 변위 량을 추정하였다. 여기서 b_i ($i=1,2,3$)는 최대의 빈도에 해당하는 화소단위의 변위량이며 w_i ($i=1,2,3$)는 이들 변위량의 빈도에 해당하며 Δ 는 Δx 와 Δy 를 의미한다.

2.2 모자이크 구성 알고리즘

가. 모자이크 구성 알고리즘

식 (5)에 따라 연속되는 프레임에서 추출된 영상의 시간에 따른 변위량 Δx 와 Δy 는 동적 모자이크 구성에 있어 워핑 파라메터로 이용된다. 시간 t에서 모자이크 M_t 는 시간 t 및 t 이전의 k개의 영상의 함수

$$M_t = f(I_t, I_{t-1}, I_{t-2}, \dots, I_{t-k}) \quad (6)$$

에 의해 표현될 수 있으며, 동적 모자이크의 경우

$$\begin{aligned} M_0 &= I_0 \\ M_t &= f(\tilde{I}_t, M_{t-1}) \end{aligned} \quad (7)$$

과 같이 순차적으로 갱신할 수 있다. 식 (7)에서 \tilde{I}_t 는 시간 t에서의 디코딩되고 Δ 만큼 워핑이 일어난 프레임에 의미한다.

만약 카메라의 움직임이 작은 경우에는 연속된 프레임의 변위가 1 화소 미만인 경우에 화소단위의 워핑을 고려하기 때문에 모자이크는 갱신되지 않는다. 그러나, 매 프레임의 변위는 1 화소를 넘지 않아 갱신이 되지 않더라도 여러 프레임이 진행되는 동안에 누적되는 변위는 1 화소를 넘어갈 경우 모자이크는 갱신되어야 마땅하다. 즉 순차적인 변위가 0.4, 0.5, 0.3 ... 식으로 진행되면 처음 두 프레임에 의한 모자이크 갱신은 일어나지 않지만 누적된 변위가 1을 넘어간 시점인 세 번째 프레임은 1 화소만큼 워핑이 일어나 모자이크 갱신에 기여해야한다. 알고리즘의 Step 2-5는 이러한 과정을 반영하고 있다.

이러한 점들을 고려하여 본 논문에서는 다음과 같은 알고리즘에 의해 모자이크를 구성하였다.

<모자이크 구성 알고리즘>

Step 1: Initialization

$M_0 = I_0$, the decoded first frame of a shot.

Accumulated $\Delta Ax=0$, $\Delta Ay=0$

Step 2: Calculate \tilde{I}_t from decoded I_t

2-1: Estimate optical flow from compressed video.

2-2: Smooth by spatiotemporal median filtering.

2-3: Estimate Δx and Δy .

2-4. Calculate the displacement

If $\Delta x < 1$ and $\Delta y < 1$,

$$\Delta Ax = \Delta Ax + \Delta x,$$

$$\Delta Ay = \Delta Ay + \Delta y$$

Endif

$$displacement_x = \lfloor \Delta Ax \rfloor$$

$$displacement_y = \lfloor \Delta Ay \rfloor$$

$$\Delta Ax = displacement_x - \lfloor \Delta Ax \rfloor$$

$$\Delta Ay = displacement_y - \lfloor \Delta Ay \rfloor$$

2-5 Take warping according to the

$displacement_x$ and $displacement_y$

to make \tilde{I}_t

Step 3: Blend and update mosaic

나 혼합(blending) 방법

혼합방법이란 모자이크를 구성할때 연속된 영상의 중첩된 영역에서 화소 값을 결정하는 방법을 의미한다. 중첩된 영역에서 화소 값을 결정하는 방법에는 중첩된 영상들의 평균 또는 메디안 등 다양한 연산을 취할 수 있다. 일반적으로 혼합방법은 추정(estimation)문제라 간주될 수 있다. 식 (6)에서 시간 t에 갱신된 모자이크의 화소값을 시간 t에서의 프레임 \tilde{I}_t 를 포함하여 \tilde{I}_{t-k} 까지의 k개의 프레임의 이동평균을 취한다거나 메디안을 대표치로 취하여 결정할 수도 있으며, Nicolas가 제안한 바와 같이

$$M_t = \alpha_t \tilde{I}_t + (1 - \alpha_t) M_{t-1} \quad (8)$$

의 선형식에서 최적의 혼합을 결정하기 위한 평가 지수를 정의하고 해석적으로 혼합에 필요한 계수 α_t 를 구할 수도 있다^[7]. 식 (8)에서 α_t 가 1에 가까운 값을 가지면 M_t 는 현재의 프레임 정보를 충실히 표현하도록 갱신되며, α_t 가 0에 가까운 값을 가지면 기존의 모자이크 정보를 유지하려 할 것이다. Nicolas가 제안한 방법에서 최적화될 평가지수는 다음과 같다.

(1)모자이크의 배경에 있어 시간적인 변화의 최소화

새로 갱신되는 모자이크 영상의 배경의 급변하면 기존의 모자이크와 경계면에 시각적인 불연속이 발생할 수 있다 이러한 불연속을 방지하기 위해서는 배경영상의 변화가 크게 나타나지 않아야 한다

(2) 모자이크 정보의 시간지연의 최소화

동적인 모자이크는 가장 최근의 영상정보를 될

수 있는 한 잘 반영하여야 한다. 즉 모자이크의 각 픽셀들은 과거의 영상들과 현재 영상의 혼합에 의해 시간지연이 발생하는데 이 지연양이 될 수 있는 한 작아야 한다

(3) 해상도의 최대화

줌이 진행되는 연속된 영상의 모자이크는 해상도가 최대가 되도록 구성되어야 한다.

본 논문에서는 줌의 경우는 고려하지 않으므로 (1)과 (2)의 조건을 최소화하는 혼합계수 α_t 를 이용한다 또한 본 논문의 알고리즘의 Step 3에서는 평균 및 메디안, Nicolas의 분석적 가중치에 의한 선형결합 방법을 모두 적용하여 비교하였다

III. 실험 및 검토

실험에서는 줌을 포함하지 않고 비교적 카메라의 움직임이 크지 않은 MPEG-2로 압축된 비디오 데이터를 사용하였다. 이러한 데이터를 찾기 위해서는 MPEG의 압축 도메인에서 직접 셋을 분할할 수 있는 알고리즘과 MPEG의 압축 도메인에서 직접 분할된 셋에서 카메라의 동작을 분석할 수 있는 알고리즘을 이용할 수 있다^[11].

그림 9는 90개의 뉴스 비디오로부터 얻어진 MPEG 프레임들 분석적인 가중치를 적용하여 혼합시킨 모자이크를 보여주고 있다 또한 그림 10은 야구장 전경을 제안된 방법에 의해 모자이크한 영상을 보여주고 있으며, 그림 11은 영화 비디오로부터 얻은 데이터로 카메라의 팬에 의한 변화 뿐 아니라 팬과 틸트에 의한 변화도 구성된 모자이크에 잘 반영되고 있음을 보여준다.

표 1은 비디오로부터 그림 9의 모자이크를 구성하는데 혼합방법들의 오차를 비교한 것으로 현재 프레임 \tilde{I}_t 와 만들어진 모자이크 M_t 사이의 변화 에러

$$MSD1 = \sum_{p \in Z} (\tilde{I}_t(p) - M_t(p))^2 \quad (9)$$

와 현재 만들어진 모자이크 이미지 M_t 와 이전 모자이크 이미지 M_{t-1} 사이의 차분 에러 관점에서

$$MSD2 = \sum_{p \in Z} (M_t(p) - M_{t-1}(p))^2 \quad (10)$$

평균, 메디안 그리고 분석적 가중치에 의한 모자이크 구성 방법의 성능 비교를 표현하고 있다 평균을

이용하는 경우에는 식 (8)에서 $\alpha=0.5$ 인 경우에 해당하며 메디안을 이용하는 경우에는 식 (6)에서 위 평된 $\hat{I}_n, \hat{I}_{n-1}, \hat{I}_{n-2}$ 의 중첩된 부분의 메디안을 취하였다. 식 (9)와 (10)에서 Z 는 두 영상의 공통부분으로, MSD1은 시간적 지연을 간접적으로 표현하며 MSD2는 구성된 모자이크의 공간적 변화를 표현한다. 표 1에서 알 수 있듯이 식 (9)와 (10)의 관점에서 최적화된 가중치를 사용하는 Nicolas의 방법이 이들 관점에서 가장 우수한 성능을 보였다. 또한 그림 8은 그림 9를 얻는 과정에서 계산된 가중치의 변화를 보여 주고 있으며, 고정된 가중치를 사용할 경우 $\alpha=0.65$ 로 선택하면 근사적으로 최적의 결과와 유사함을 알 수 있다.

또한 실험에서 식 (3)의 기하학적인 변환 모형이 카메라가 배경으로부터 멀리 떨어진 경우, 카메라의 팬 동작만 있는 경우와 같이 파노라마 구성시 각 프레임의 중앙 부분에서 수평축의 변위로 최신의 근사화가 될 수 있기 때문에 프레임의 중앙에서만 혼합영역을 고려할 수도 있다.

표 1. 각 방법의 MSD1과 MSD2 누적 에러

	MSD1	MSD2
Nicolas 방법	5,829,588	5,218,369
평균을 이용한 방법	6,312,832	5,666,930
메디안을 이용한 방법	6,241,180	5,582,521

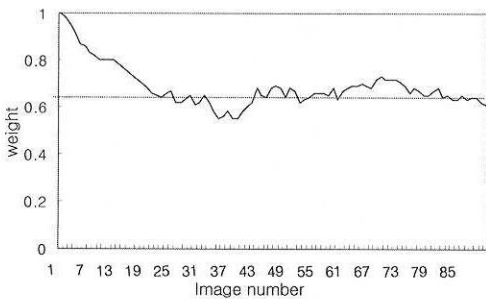


그림 8. 분석에 의한 가중치 변화

IV. 결론

본 논문에서는 MPEG2 비디오 스트림으로부터 복호화 과정 없이 매크로 블록 단위의 광 플로우를 근사적으로 구성하고, 단순한 카메라 움직임에 기인한 변위량을 추출하여 이를 모자이크 합성에 이용

할 수 있음을 보였다. 구성된 모자이크는 광 플로우는 매크로 블록단위로 부호화 과정에서 포함된 움직임 벡터로부터 간단하게 구성하였으며, 이러한 근사적인 광 플로우로부터 Hough 변환을 이용하여 카메라의 팬 및 틸트, 줌에 따른 연속되는 프레임의 변위를 추정하였다. 제안된 방법의 Hough 변환은 어큐멀레이터가 1차원적으로 구성되기 때문에 계산량과 메모리 요구량은 크지 않으며, 평균지승 오차에 의한 추정방법에 비해 강인하고, outlier를 배제하기 위한 전처리 과정이 불필요하다. 모자이크 이미지 구성을 위해서는 Nicolas가 제안한 분석적 가중치를 이용하는 혼합 방법과 평균 또는 메디안 연산 등을 검토하였는데 Nicolas가 제안한 분석적 가중치를 이용하는 혼합 방법이 평균 또는 메디안 연산보다 우수함을 보였다.

그러나, 제안된 방법에서 광 플로우는 매크로 블록 단위로 근사화된 형태이기 때문에 줌에 대한 모자이크를 구성하는 데는 제약이 있었으며, 수평 및 수직축의 변위에 대한 모자이크만을 고려하였다. 제안된 방식은 모자이크를 셋 단위로 구성하여 비디오 검색에서 대표 프레임으로 이용할 수 있다. 한편 본 논문의 모자이크는 압축비디오의 정보를 이용하기 때문에 빠르게 구성될 수 있다는 장점이 있는 반면, 제한된 정보를 이용하기 때문에 비디오 압축이나 편집 등과 같은 용도의 정확한 모자이크 구성에는 적절하지 않다.



(a) 3 frame

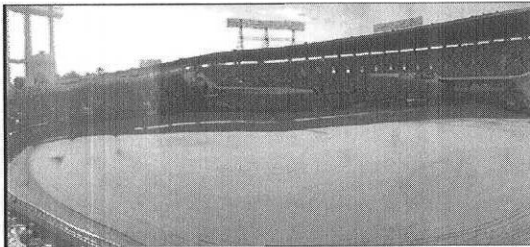


(b) 70 frame

그림 9. 분석적 가중치에 의한 동적 모자이크



(a) 3 frame



(b) 190 frame

그림 10. 제안된 방법에 의한 모자이크



(a) 3 frame



(b) 80 frame

그림 11. 팬과 틸트가 동시에 일어난 동영상의 모자이크

참 고 문 헌

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

[2] Irani, M and Anandan, p. and Hsu, S. "Mosaic based representations of video sequences and their applications," *IEEE Computer Vision*, pp.605-611, 1995.

[3] V. Kobla, and D. Doermann "Compressed domain video indexing techniques using DCT and motion vector information in MPEG video." *In Proc. Of SPIE*, vol.3022, pp.200-211, 1997.

[4] J. Meng, and S.-F. Chang, "Tools for Compressed domain Video Indexing and Editing," *SPIE Storage and Retrieval for Still Image and Video Databases*, Vol. 2670, pp.180-191, 1996.

[5] R. Milanese, F.Deguillaume, A. Jacot-Descombes "Efficient Segmentation and Camera Motion Indexing of Compressed Video," *Nortim*. 1998. 01.

[6] 유원영, 최정일, 이준환, "일반화된 Hough 변환기법을 이용한 MPEG2 압축영역에서의 카메라의 움직임 해석," *한국정보처리학회 논문지*, 제7권 제11호, pp. 3566-3575, 2000.

[7] H. Nicolas "New methods for dynamic mosaicking", *IEEE Trans. Image Processing*, vol. 10, pp.1239-1251, Aug. 2001.

[8] Isaac Cohen, and Gerard Medioni, "Detection and Tracking of Objects in Airborne Video Imagery," *CVPR'98 Workshop on Interpretation of Visual Motion*, 1998.

[9] ISO-IEC 1-1/ISO-IEC 13812-2 International Standards, 1st Ed., 1996.

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

[11] 김성국, 박영규, 유원영, 김준철, 이준환, "MPEG-2 압축영역의 TV 스포츠 뉴스 색인은 위한 장면전환 및 기사검출", *한국 정보처리학회 논문지*, 제6권 제6호, pp1703-1712, 1999.

설 정 규(Jung-kyu Seol)



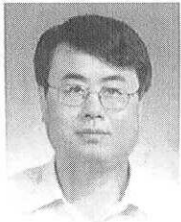
1998년 2월 : 우석대학교
물리학과 졸업(학사)
2003년 1월 : 전북대학교
전자공학과 재학(석사)
<주관심 분야> 영상처리,
인공지능

이 승 희(Seung-hee Lee)



2002년 2월 : 전북대학교
전자공학과 졸업(학사)
2003년 1월 : 전북대학교
전자공학과 재학(석사)
<주관심 분야>멀티미디어,
영상처리

이 준 환(Joonwhoan Lee)



1980년 2월 : 한양대학교
전자공학과 졸업(학사)
1982년 2월 : 과학기술원
전기전자공학과 졸업
(석사)
1990년 : 미국 미주리대학
컴퓨터공학과 졸업(박사)

<주관심 분야>영상처리, 컴퓨터 비전, 인공지능