2001年度 第14回 信號處理合同學術大會 論文集 第14卷 1號

Mellin 변환을 이용한 동영상 시퀀스의 모자이크 기반 표현

장영준, 하경민, 박준희*, 이병욱*, 최윤식
연세대학교 전기전자연구실
*이화여자대학교 정보통신학과

Mosaic based representations of video sequences using Mellin transform

Young June Jang, Kyung Min Ha, Jun Hee Park*, Byung Uk Lee*, Yoonsik Choe
Department of Electrical and Electronic Engineering, Yonsei University
Department of Information Electronics, Ewha woman's University
E-mail : lupin@image3.yonsei.ac.kr

요 약

동영상은 각 프레임 사이에 시간적으로나 공간적으로 많은 양의 정보가 중복되어 있다. 이러한 중복 정보를 줄이는 표현 방법들 중에 하나로, 동영상을 커다란 하나의 영상으로 정합하여 중복 정보를 줄이는 모자이크 기법이 있다. 두 개 이상의 영상을 정합하기 위해서는 영상간의 카메라 파라미터가 필요한데, 본 논문에서는 Mellin 변환을 사용하여 카메라 파라미터를 구하였다. 이때 3차원 공간모델은 직교 투사법을 사용하였으며, 영상의 움직임 모델로는4개의 파라미터(평행이동, 확대/축소, 회전)를 사용한 어파인 움직임 모델을 사용하였다. 이렇게 구현된 파노라마 영상은 동영상에서 움직이는 물체를 검출하거나 추적하고, 동영상을 편집하는데 응용될 수 있다. 또한 본 연구의 최종 목적인 3D 영상의 배경을 구현하는 데 좀 더 사실적인 영상을 제공할 수 있다.

1. 서 론

동영상은 다음의 2가지 관점에서 정지영상에 비해 장점을 가진다; (i) 시간적인 정보를 표현할 수 있다. (ii) 임의의 장면들에서 연속적으로 변화하는 정보를 표현할 수 있다. 그러나, 이러한 점은 반대로 각 프레임(frame)간에 중복정보(redundancy)를 포함하고 있기 때문에 전체적인 정보의 양이 증가하는 것을 의미한다.

본 논문에서는 2개 이상의 영상으로부터 2차원 카메라 파라미터(평행이동, 확대/축소, 회전)를 구하고 정합(registration)하는 모자이크 기법에 대해서 알아보고자 한다. 모자이크 기법은 여러 장의 동영상을 하나의 커다란 정지영상(panorama)으로 표현함으로써 각 프레임간에 존재하는 중복정보를 제거하는 방법이다.[1][2]각 프레임간의 카메라 파라미터는 Mellin 변환을 사용하여 구한다.[3] Mellin 변환은 영상을 Fourier 변환하여 위상을 비교하고, 이를 역 Fourier 변환하고 최대치를 찾는 방법으로 영상간의 평행이동을 구하는 방법이다. 이러한 방법을 응용하여 회전각도는 회전방향으로 양자화하여 추출하고, 크기를 log -polar scale로 변환하고 양자화하여 확대/축소 파라미터를 추출할 수 있다. 이렇게 구한 각 프레임간의 카메라 파라미터를 사용하여 직교투사법(orthogonal prjection)으로 가정하에 하나의 기준프레임에 모자이크되어 파노라마 영상을 만들게 된다.

이렇게 만들어진 파노라마 영상은 영상내에 움직이는 물체를 검출하고, 추적하는 데에도 사용된다.[4] 즉, 움직이는 물체가 각 프레임에 따로따로 존재하는 것이 아니라, 하나의 커다란 영상에 표현될 수 있기 때문에 움직이는 물체의 경로를 쉽게 추정할 수 있다. 이 외에도 최근에 관심의 초점이 되는 동영상의 편집이나 영상의

데이터베이스에도 응용될 수 있다.[2]

2. 모자이크 기반 표현

2.1 모자이크 표현법

모자이크 방법은 동영상 시퀀스의 모든 프레임을 하나의 영상에 정합하여, 프레임에 한정된 영상을 확장된 파노라마 영상으로 만드는 방법이다. 모자이크 영상은 크게 움직임 물체를 무시한 static 모자이크와 움직임 물체를 고려하여 최종적인 파노라마에 각 프레임의 움직임 물체가 모두 표시되는 dynamic 모자이크로 나뉜다. Dynamic 모자이크는 움직임 물체의 검출이나 추적, 영상정보의 편집 등이 유리하기 때문에 본 실험에서는 dynamic 모자이크를 사용하였다.

2.2 카메라 움직임의 모델링

(X_p,Y_p,Z_p)를 3차원의 좌표계라 하고, (x,y)를 영상의 좌표계라고 하면, 본 논문에서는 직교투사법으로 근사화하였으므로, 다음과 같은 관계가 있다.

$$x_i = X_i \qquad y_i = Y_i \,, \qquad i=\{1,2.....\} \tag{1}$$

여기서 i는 시간적으로 앞,뒤 프레임을 나타낸다. 본 논문에서는 카메라에서 물체까지의 거리가 먼 영상으로 파노라마 영상을 만드는 것을 목표로 하였기에 직교투사법으로 근사화 하여도 무방하였다. 각각의 프레임간의 관계는 다음과 같은 어파인(affine) 모델로 근사화하였다.

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = s\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \tag{2}$$

여기서 $s=Z_1/Z_2$로서 확대/축소 성분, θ는 각 프레임간의 회전성분이고, (x_0,y_0)는 영상의 평행이동 성분이다.

3. 알고리듬

3.1 Mellin 변환

기본적인 Mellin 변환은 Fourier 변환의 shift theorem을 이용한 것이다. 영상 $f_1(x,y)$와 $f_2(x,y)$는 (x_0,y_0)의 평행이동 성분을 갖는다고 가정하자. 즉,

$$f_2(x,y) = f_1(x-x_0, y-y_0) \tag{3}$$

양변에 Fourier 변환을 취하면,

$$F_2(\omega_x,\omega_y) = e^{-j2\pi(\omega_x x_0+\omega_y y_0)} F_1(\omega_x,\omega_y) \tag{4}$$

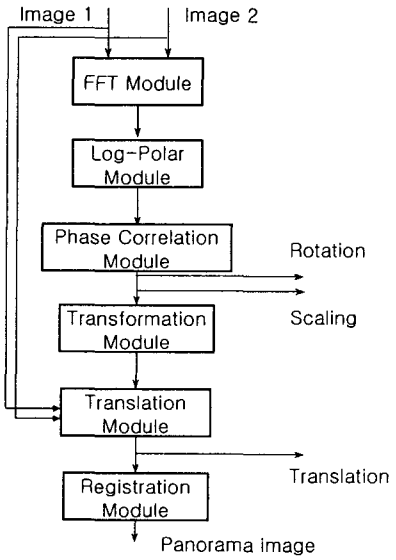


그림 1. Mellin 변환의 블록 다이어그램

최종적으로, 두 영상의 cross-power spectrum은 다음과 같다.

$$\frac{F_1(\omega_x,\omega_y)F_2^*(\omega_x,\omega_y)}{|F_1(\omega_x,\omega_y)F_2(\omega_x,\omega_y)|} = e^{-j2\pi(\omega_x x_0+\omega_y y_0)} \tag{5}$$

여기서 F_2^*는 F_2의 complex conjugate이다. 주파수 영역에서 역 Fourier 변환을 하게 되면, 두 영상간의 차이를 impulse 함수의 형태로 얻을 수 있다. 즉, 이것이 두 영상 사이의 평행이동 성분이다. 회전성분이나, 확대/축소 성분은 약간의 전처리 과정을 필요로 한다. 원영상들을 log-polar좌표계로 변환하고, 양자화한 후에 앞에서 언급한 phase correlation theorem을 이용하여 성분들을 구하게 된다.

그림 1은 본 실험의 전체적인 알고리듬을 보여준다. 그림에서처럼, 모든 성분의 계산은 Mellin 변환을 기본으로 한다.

3.2 영상 정합

파노라마 영상을 만들기 위한 영상정합의 방법에는 다음과 같은 대략 4가지 방법이 있다.

i) 보정된 영상들의 화소값의 평균을 이용

ii) 보정된 영상들의 화소값의 median을 이용

iii) 보정된 영상들의 화소값에 가중치(weight)를 부여한 평균이나, median을 이용

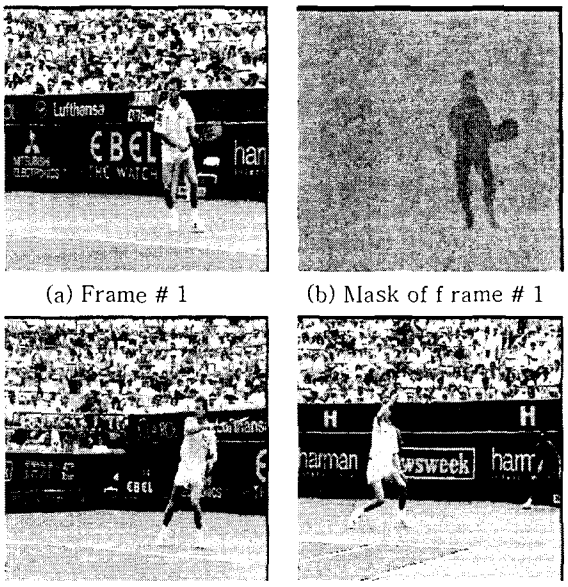
iv) 가장 최근의 프레임으로 update

여기서 i)과 ii)의 방법은 움직이지 않는 배경 영상은 정확하게 나오나, 프레임 내에서 움직이는 물체가 존재하면 움직이는 물체가 blur되어 사라지는 경향이 있으나, iii)과 iv)의 방법은 움직이는 물체를 파노라마 영상에 표현해 줄 수 있는 장점이 있다. 일반적으로 iv)의 방법이 가장 보편적인 방법이므로, 본 실험에서는 iv)의 방법을 이용하여 영상들을 정합하였다. 이때, 전경과 배경이 분리된 마스크를 사용하여 전경이 없어진 최종적인 파노라마 영상을 만들어 보았다.

그리고, Mellin 변환을 이용하면 사전에 제공된 마스크 없이도 배경과 다른 속도로 움직이는 전경을 배경으로부터 분리하거나 추적할 수 있다.[5] 따라서 본 실험에서는 동영상 시퀀스와 Mellin 변환만을 이용하여 특별한 영상분할 알고리듬 없이 전경과 배경을 분리하는 마스크를 만들어 보았으며, 이 마스크가 정확하고 빠른 영상 정합에 적합한지도 알아보았다.

4. 실험결과

실험에 사용한 영상은 MPEG-4의 실험영상으로 많이 사용되는 "Stefan" 시퀀스를 사용하였다.(256×256 byte) 이 영상의 예가 그림 2에 나와있다. 이 영상은 각각 다르게 움직이는 테니스 선수의 전경(foreground)과 관중의 배경(background)으로 확실히 구분할 수 있으며, 특별히 그림 2.(b)는 전경과 배경을 분리하는 마스크의 한 예이다. 이 시퀀스는 큰 x축 평행이동성분과 약간의 y축 평행이동성분, 그리고 약간의 확대/축소 성분이 존재하는 시퀀스이다.



(a) Frame # 1 (b) Mask of f rame # 1

(c) Frame # 120 (d) Frame # 200

그림 2. "Stefan" sequence

그림 2의 실험영상을 사용하여 파노라마된 영상의 결과가 그림 3에 나와있다. 파노라마 영상으로 모자이크 할 때, 새로 들어온 입력 영상들을 update하는 방식으로 모자이크 하였으며, 앞에서 언급한대로 전경과 배경 마스크를 사용하여 파노라마 영상으로 모자이크 하였다.

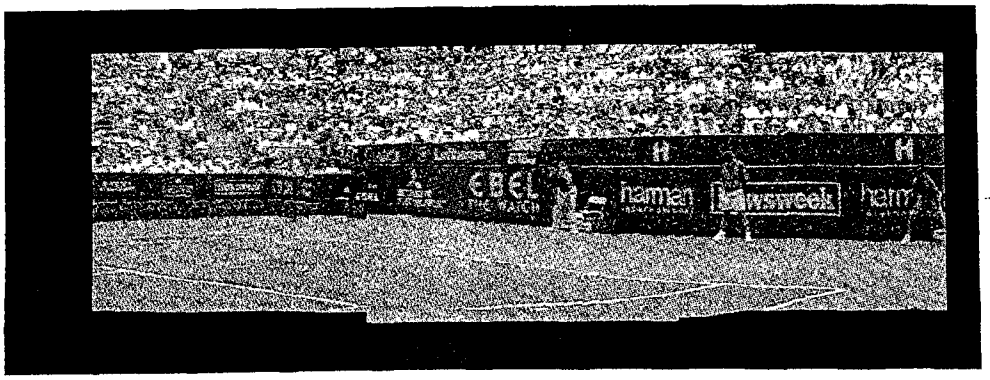


그림 3. 파노라마 영상

따라서, 파노라마 영상에는 전경인 테니스 선수가 보이지 않는다. 하지만, 시퀀스의 영상이 차례로 정합되면서 움직임 파라미터들의 오차가 누적됨에 따라 약간의 에러가 발생함을 알 수 있다. 이것은 전체 시퀀스를 조금씩 정합하고, 이렇게 정합된 작은 파노라마 영상을 또 다시 정합하거나, 좀 더 정확한 카메라 모델링을 사용함으로써 해결 가능하리라 생각한다.

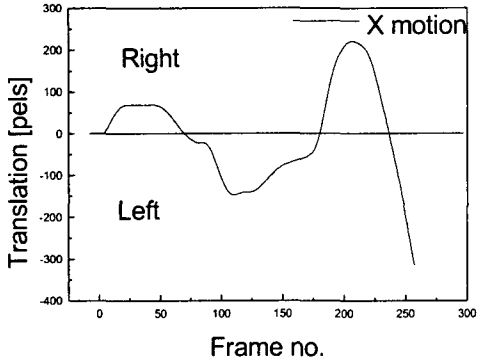


그림 4. X축 평행이동 성분

표 1. Global motion과 local motion의 비교

motion / Frame no.	Global motion (배경)	Local motion (전경)
# 1 - # 2	(3,0)	(0,0)
# 2 - # 3	(7,0)	(7,0)
# 3 - # 4	(12,0)	(12,0)
# 4 - # 5	(17,0)	(15,0)
# 5 - # 6	(23,0)	(19,0)
# 6 - # 7	(29,0)	(22,0)
# 7 - # 8	(35,-1)	(35,-1)
# 8 - # 9	(40,-2)	(39,-2)
# 9 - # 10	(40,-3)	(31,-3)

그림 4는 시퀀스의 가장 주된 움직임인 x축 평행이동 성분만을 따로 추출한 결과이다. 그리고 표 1은 Mellin 변환으로 추출한 카메라의 global motion(배경)과 움직임 개체의 local motion(전경)을 10 프레임까지 시퀀스별로 추적한 결과이다. 이 결과를 이용하면, 특별한 영상분할 알고리듬 없이 시퀀스에서 움직임 개체(전경)을 추출할 수 있으며, 좀 더 정확한 영상정합을 할 수 있다. 이 결과가 그림 5에 나와 있다.

5. 결 론

본 연구는 3차원 영상의 배경으로 쓰일 수 있는 현실감 있는 2차원 배경의 구현을 목표로 시작되었다. 동영상 시퀀스를 바탕으로 모자이크 기법을 사용하여, 파노라마 영상을 만들어 보았다. 각 프레임 사이의 움직임 성분은 Mellin 변환을 사용하여, 빠른 시간에 정확하게 구하였으며, 이를 바탕으로 시야각이 넓은 파노라마 영상을 만들어 보았다. 그리고 특별한 영상분할 알고리듬없이 배경과 다르게 움직이는 전경을 분할할 수 있었으며 이를 바탕으로 좀 더 정확한 영상정합을 할 수 있었다.

6. 참고문헌

[1] M.Irani, P.Anandan, and S.Hsu "Mosaic Based Representations of Video Sequences and Their Applications", *Computer Vision Proceedings*, Fifth International Conference on 1995, pp. 605 - 611, 1995

[2] M.Irani and P.Anandan "Video Indexing Based on Mosaic Representations", *Proceedings of the IEEE*, vol. 86, no. 5, pp. 905-921, 1998

[3] B.Srinivasa Reddy, and B.N.Chatterji "An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration", *IEEE Trans. Image Processing*, vol. 5, no. 8, pp. 1266-1271, Aug. 1996

[4] J.Davis "Mosaics of Scenes with Moving Objects" *Computer Vision and Pattern Recognition Proceedings*, pp. 354-360, 1998

[5] 박수현, 이병욱 " Mellin transform에서의 다중 물체 이동 검출", *신호처리합동학술대회*, pp 199-202, 1999