

다중 영상에 의한 비디오 모자이크 시스템

양 원 보* · 임 문 순** · 이 양 원*

*군산대학교 컴퓨터과학과

**충남 시초 초등학교

Video Mosaic System by Multi-Image

Wonbo Yang* · Moonsun Lim** · Yangweon Rhee*

*Dept. of computer Science, Kunsan National University

**Sicho Elementary school , Choongnam

E-mail : bbong@ai.kunsan.ac.kr

요약

단편적인 의미를 지니는 각각의 영상들보다 이들 영상의 내용을 함축적으로 내포하는 하나의 영상으로 표현하는 것이 많은 효과를 제시하고 있다. 이 상황에 적용되는 기술이 '모자이크'이다. '모자이크'라고 하는 것은 여러 영상들 중에서 중복되는 영역들을 제거하여 각각의 영상들을 하나의 복합된 새로운 영상으로 생성하는 것이다.

본 연구는 다 영상에 의한 모자이크 시스템을 개발한다. 시스템은 장면분할, 모자이크 영상의 생성으로 구분된다. 장면분할에서는 동영상인 비디오 데이터들을 연속된 시간 내에서 유사한 영상들끼리의 묶음으로 구분하고, 모자이크 영상의 생성은 분할된 장면 안의 모든 프레임들 한 영상으로 구성한다

ABSTRACT

It is presented many effect that represented by implicated one image more than each images with a fragment meaning. 'Mosaic' is technique be applied at this situation. 'Mosaic' is created by complicated one new image which by multi image be eliminated overlap region.

This research is development for mosaic system by multi image. The system is divided that shot segmentation and mosaic image creation. Shot segmentation divided that merge with respect to similarity images which video data of moving picture in sequence time and mosaic image creation is composed of one image with which all frames in segmented shot.

I. 서론

최근에는 멀티미디어 기술들이 발전하고 널리 보급되면서 방송, 의료, 교통 등 생활 전반에 걸쳐 그 수요가 크게 증가하고 있다[13]. 더불어, 이용의 폭도 넓어져 멀티미디어 자료의 함축적인 의미 정보의 활용에 적극적인 요구가 제시되고 있는 실정이다. 멀티미디어 기술의 분야는 영상처

리와 멀티미디어 데이터베이스, 컴퓨터 통신의 결합된 통합 기술로 구성되어진다. 통신과 멀티미디어 기술의 발달로 인하여 대용량의 멀티미디어 데이터의 사용이 실현되었고 중요한 정보의 교환 수단으로 발전하고 있으며 멀티미디어 기술의 핵심 분야로 취급되고 있다. 그러나 멀티미디어 데이터는 중요한 분야임에도 불구하고 그 용량이나 복잡성 때문에 다루기가 쉽지 않다. 특히 비디오

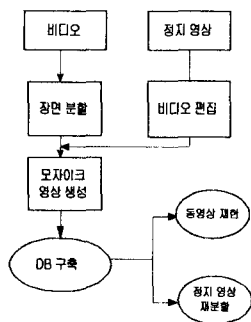
데이터는 그 용량에서부터 엄청난 크기를 가지기 때문에 효율적으로 저장, 관리하고, 사용자가 쉽게 접근할 수 있는 방법이 요구되고 있다[13,14].

본 논문에서의 모자이크 시스템은, 이러한 사용자들의 필요성으로부터 시작한 것으로 나누어진 영상을 일치하는 부분을 찾아서 하나의 영상으로 합하고자 하는 것이다. 이 시스템은 연속된 비디오를 통해서 영상들을 입력할 수도 있고 또한 휴대용(hand-handle) 카메라, 혹은 유사한 영상들을 가지고서 입력하여, 사용자의 상호작용을 통한 능률적인 모자이크 구현에 대한 시스템이다. 사용자의 상호작용이란, 모자이크 영상 구축 시, 사용자가 '비디오 편집'을 통해서 직접 원하는 영상을 고를 수 있게 하는 것을 말하고 또한, 사용자가 원하는 모자이크 표현을 선택할 수도 있게 하는 것이다.

모자이크 영상은 연속된 비디오에 포함된 많은 정보들을 표현해 줄 수가 있다. 모자이크를 표현하는데 있어서는 정적(static), 동적(dynamic), synopsis 모자이크로 분류할 수 있다[1][2]. 모자이크 구축 과정의 기술들은 연속된 영상의 정렬, 통합, 중복성에 대한 분석으로 이루어진다. 비디오 압축, 확장된 시각화, 비디오 향상 등을 포함하는 비디오 인덱싱, 탐색, 조정에서 모자이크 표현은 강력한 비디오 응용기술이다.

모자이크 시스템에 대한 응용분야로는 상당히 많이 존재하는데, 그 중에 가장 전통적인 분야는 공중 인공위성 사진의 구축이고, 최근의 연구 분야는 장면 고정과 change detection, 비디오 압축과 인덱싱, 심지어 간단한 사진 편집까지도 연구되고 있다.

II. 본론



[그림 1] 시스템 개략도

(그림 1)은 입력 단계에 있어서 카메라로부터 찍은 비디오, 혹은 핸드-핸들 카메라(hand-handle camera)로부터 찍은 정지 영상을 가진다. 입력된 영상이 카메라로부터 찍은 일련의 비디오 영상이라면 위에서 이야기한 장면 분할의 방법을 거치고 하나의 영상으로 합성하는 모자이크 기술의 과정은 비디오나 정지영상 모두 이루어지는 작업

이다. 하지만, 비디오와 정지 영상은 각 특성에 따라 사용자의 상호작용을 통해 더욱 빠른 모자이크를 구축할 수 있게 된다. 본 연구에서는 이러한 과정에서 비디오에 대한 장면 분할과 정지 영상을 처리하기 위한 편집 기술, 그리고 그러한 비디오와 정지 영상을 가지고 모자이크 영상을 생성하는 것에 중점을 두고 있다.

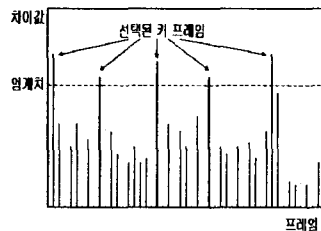
2.1. 비디오 분할(장면 전환) 기법

'모자이크'라고 하는 것이 중복되는 영상들을 하나의 영상으로 만들겠다는 것이기 때문에 합치려는 영상들이 같은 내용들로서 이루어져 있어야 한다. 연속된 비디오에서 이런 같은 의미들을 지닌 것을 장면이라고 부르는데 이 장면을 먼저 분할해내야 한다. 내용을 추출하기 위해서는 우선 장면 분할 작업이 선행되어야 한다. 장면을 분할하는 방법에는 여러 가지가 있다. 장면을 분할하기 위한 방법에는 여러 가지가 있다. 프레임간의 차이를 이용하는 방법과 칼라 히스토그램을 이용하는 방법, 객체의 이동을 계산하여 사용하는 방법, 윤곽선(edge)을 추출하여 이용하는 방법과 전체적인 웨이블릿(wavelet)을 측정하여 이용하는 방법 등 많은 방법이 사용되었다. 이러한 방법 중 주로 히스토그램 차이를 계산하여 사용하는 방법이 많이 사용되었는데, 크게 칼라 히스토그램과 흑지 히스토그램이 사용되었다. [14]

(그림 2)는 일반적인 히스토그램 차이 값을 적용시키는 방법에 대한 그림이다. 사용된 식은 다음과 같다.

$$S(i) = \sum_{j=0}^{N-1} |H_i(j) - H_{i+1}(j)| \quad (1)$$

$S(i)$ 는 이웃 프레임과의 유사성을 측정하는 함수로, $H_i(j)$ 는 i 번째 프레임의 j 번째 히스토그램 요소 값이라는 의미이다. 첫 프레임과 끝 프레임의 두 번째 프레임으로 선택한다. 혹은, 각 장면의 두 번째 프레임으로 선택하는 간단한 방법을 들 수 있다. 하지만, 그 프레임이 한 장면의 많은 정보를 포함하지 못 할 수 있기에, 다음 장에서 모자이크 영상 생성



[그림 2] 히스토그램 차이 값에 의한 키 프레임 선택

위의 방법은 일반적인 방법이므로 데이터의 성

격에 따라 여러 가지 방법을 적용하여 히스토그램 차이를 구할 수 있다.

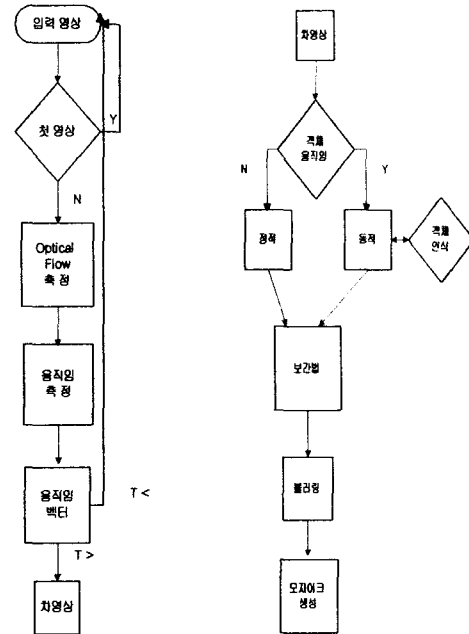
III. 모자이크 기술

[표 1] 모자이크의 장점 및 필요성

모 자 이 크	장점	<ul style="list-style-type: none"> • 시각적 유리성 • 압축 효과 (저장) • 쉬운 인덱싱 • 객체의 움직임 인식
	필요성	<ul style="list-style-type: none"> • 고행상도의 영상 요구 • 가상 현실 • 먼 거리 여행 • 위성 사진 및 지도제작

위의 (표1)는 모자이크의 장점 및 필요성을 간략하게 표로 나타낸 것이다. 모자이크의 장점과 활용 적인 측면을 보면, 여러 영상을 하나의 영상으로 나타내 줌으로써, 비디오의 연속된 영상 속에서 많은 정보들을 한번에 볼 수 있다는 장점을 첫 번째로 꼽을 수 있으며, 또한 하나의 영상으로 만들기 때문에 반대로의 과정으로 영상을 인덱싱할 경우에 있어서, 하나의 모자이크 영상을 각각의 영상으로 정합할 때에 각 영상의 위치 값을 저장시켰다가 그 부분만을 잘라내 주면 되기에 저장적인 측면과 인덱싱하는 과정에 더욱 효율적이다[1]. 또한 Synopsis 모자이크를 통해서 객체의 움직임의 과정을 볼 수 있게된다. 즉, 예를 들면 비행기의 궤도를 몇 개의 영상을 모자이크화해서 보면, 그 행로를 알 수가 있게 되는 것이다. 또한, 운동경기에 있어서도 응용을 들 수가 있다. 축구 경기장에서 각 선수와 공의 움직임을 한 눈에 알 수 있게 허락해 준다.

(그림 3)은 모자이크 기술에 관한 부분을 더욱 세부적으로 나타내 준 것으로 모자이크 영상 생성을 위한 시스템 흐름도를 개략적으로 표현한 것이다. 간략하게 살펴보면, 카메라로부터 찍혀진 실세계의 영상들을 받아들여서 optical flow를 계산한다. 여기서, 중요한 작업은 배경과 객체에 대한 구분이 필요한데, 이 과정에서 우리는 일반적으로 전통적인 사고를 가질 수 있게된다. 카메라가 이동할 때 영상의 모서리 부분이 배경이라는 점과, 객체의 움직임은 중심에 놓이게 된다는 것이다. 이렇게 motion 측정을 한 후에 motion vector를 계산하게 된다. 계산된 motion vector를 가지고 총체적(global), 지역적(local) 탐색(search)을 하여 차 영상, 즉 t 영상과 t-1 영상의 차이 값을 계산하게 된다. 두 영상에서 움직임 벡터(motion vector)의 값이 기준 값보다 작게 되면 이어지는 다음 영상을 받아들리게 된다.



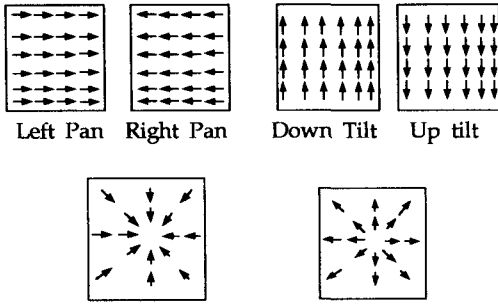
[그림 3] 모자이크 영상 생성을 위한 시스템 흐름도

이런 작업들을 거친 후에 보간법(interpolation)의 과정을 거쳐 각 영상들을 합치게 되고 흐리게 하기(blurring)의 과정을 통해 각 합쳐진 부분에 대해서 영상을 완만하게 한다. 위의 총체적인 작업들을 거친 후에 모자이크를 생성하게 된다.

모자이크 기술의 기본적인 사고는 연속된 완전한 비디오에서 분할된 각 장면을 가지고 하나의 영상을 생성하는 것이다. 그러므로, 하나의 영상으로 합쳐진 영상은 장면의 모든 정보를 포함하고 있다.

3.1 Optical Flow를 이용한 Motion Vector 측정

모자이크를 구축하기 위해서, 카메라의 움직임을 측정하는 것이 필요한데, 이는 optical flow를 이용하여 계산되어 질 수 있다[4]. 아래 (그림 4)가 일반적인 카메라의 움직임을 구분해 놓은 것이다. 카메라가 왼쪽으로 이동(left pan)이 되면 영상내의 객체는 오른쪽으로 이동한 것이 되며, 카메라의 오른쪽 이동(right pan)은 객체의 움직임이 왼쪽으로 이동되어지는 것이다. 또한, 카메라가 아래로 움직(down tilt)이면 객체는 위로, 카메라가 위로 움직(up tilt)이면 객체는 아래로 이동하는 것이 되어진다. 마찬가지로 신축(scale)에 있어서 카메라를 zoom out 시키면 객체는 영상 내로 이동하는 것이 되어지며, zoom in이 되면 객체는 영상 내에서 바깥쪽으로 커지게 되어진다.



[그림 4] pan, tilt and zoom에 의해 생성된 optical flow field

3.2 사각형 차이 값의 합계(sum of squared difference) 오류 측정

$$E(\{u\}) = \sum_x (I(x, t) - I(x - u(x), t - 1))^2 \quad (2)$$

X는 x,y의 픽셀 위치이며, u는 방향벡터, t는 시간을 나타낸다, 즉, 픽셀의 x,y 위치한 t시간의 영상의 명암 값에서 t-1시간의 영상에서 픽셀의 x,y 위치의 방향벡터(x축, y축)만큼 이동한 명암 값의 차이 값을 가지고 제곱을 하여 최소 오류 값을 찾는 것이다. 이런 최소 사각형 오류 측정값을 가지고서 영상에서 객체의 움직임을 찾을 수 있게 된다.

3.3 좌표 변환

[표 2] 영상 좌표 변환 모델들

모델	X에서 X'로의 좌표 변환	파라미터
이동	$X' = X + b$	$b \in \mathbb{R}^{2 \times 1}$
Affine	$X' = AX + b$	$A \in \mathbb{R}^{2 \times 2}, b \in \mathbb{R}^{2 \times 1}$
양선형	$x' = q_{x'xy}XY + q_{x'xx}X + q_{x'yy}Y + q_{x'}$ $y' = q_{y'xy}XY + q_{y'xx}X + q_{y'yy}Y + q_{y'}$	$q_i \in \mathbb{R}$
투상	$X' = \frac{AX + b}{C^T X + 1}$	$A \in \mathbb{R}^{2 \times 2}, b, C \in \mathbb{R}^2$
Pseudo-perspective	$x' = q_{x'xx}X + q_{x'yy}Y + q_{x'x} + q_{x'xy}XY$ $y' = q_{y'xx}X + q_{y'yy}Y + q_{y'y} + q_{y'xy}XY$	$q_i \in \mathbb{R}$

t와 t'=t+1 시간에서 두 영상간에 비교를 고려할 수 있다. 좌표변환은 $X = [x,y]^T$ 로부터 좌표의 새 집합 $X' = [x',y']^T$ 로의 정해지는 것이다. 연관된 좌표를 발견하기 위한 접근은 변환 모델의 가정에 기반을 두고 있다. 가장 일반적인 모델은 표2에서 나타나 있다.

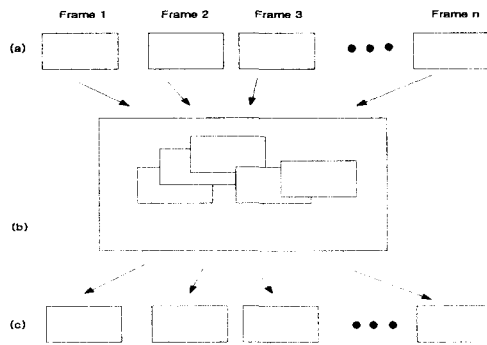
움직임 측정에서 가장 일반적인 가정이 영상이 단지 이동되었다는 것이다[3]. 이동이 된 영상은 구현하기가 쉽지만, 카메라의 회전,panning,

tilting의 여러 가지 변화에는 제어를 하지 못한다. 그래서, 본 논문에서는 또 다른 기술이 Affine인 변환을 이용한다. Affine 모델은 신축, 회전, 이동, 그리고 이것들의 결합을 포함한다. Affine 변형은 평행선들을 유지하고, 선을 따라 등간격의 점들이 또한 유지된다.

IV. 영상 재분할

모자이크 영상을 다시 각각의 영상으로 재분할 해주게 되면, 데이터의 저장 측면에 있어서 상당히 효율적으로 관리 할 수 있다. 그래서, 본 논문에서는 모자이크 영상 내에서 각 영상들과 객체의 위치를 저장시킨 후, 시간적 순서에 따라서 각 영상을 분배해 주는 방법에 대해 제안한다.

4.1 영상 재분할을 위한 데이터베이스 및 인덱싱

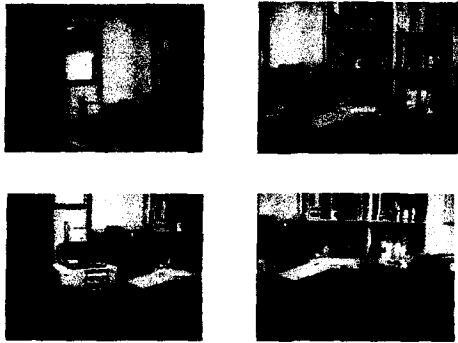


[그림 5] 모자이크 및 재분할

(그림 5)는 모자이크 영상을 만드는 것과 다시 분할하는 것을 보여주는 것이다. (a)영상은 비디오로부터 입력받은 각각의 프레임들로서 한 장면의 모든 프레임들이다. 이로부터 만들어진 모자이크 영상이 (b)영상이다. 이렇게 만들어진 모자이크 영상을 가지고서 각각의 프레임의 (x,y)위치 정보를 저장하고, 시간적인 순서에 영상들을 재분할하는 것이다. 이로 인하여 저장의 효율성과 전송시의 시간적 단축이 가능해진다. 아직까지 남은 문제점은 동적 모자이크, 즉 객체의 움직임이 있는 영상의 경우에는, 영상내의 객체의 위치 정보 및 객체의 이동 예상도 필요하게 된다.

IV. 결론

본 논문에서 제안한 모자이크 시스템을 구현하기 위해서 SONY사의 HITACHI Digital Zoom 카메라로부터 영상을 입력받았다. 실험에 사용된 영상은 320*240의 크기의 color Image를 입력 영상으로 사용하였고, 컴퓨터는 Pentium-166을 이용하였다. 또한 정지 영상의 입력을 위해 여러 장의 사진도 사용하였다.



[그림 6] 카메라로부터 얻어진 영상들



[그림 7] 모자이크 영상

위의 (그림 6)은 카메라로부터 찍은 연구실 내부의 영상들 중에서 임의의 4개의 영상을 보여준 것이다. 이 일련의 영상들을 가지고 구축된 모자이크는 (그림 7)과 같다.

향후 연구방향은 영상 재분할 시 동적 영상, 즉 객체의 움직임이 현저한 경우에 각 객체의 위치 및 움직임을 파악하고 또한 미리 예상하는 효율적인 방안에 대한 연구가 더욱 필요하며, 정지 영상의 입력 시, 사용자의 요구에 따라 영상의 삽입, 삭제에 용이하게 이루어질 수 있도록 편집 기술이 확장 되어야 한다. 그리고, 현재 트리-기반 모자이크에 대한 연구가 진행중이다.

참고문헌

[1] Michal Irani, P.Anadan, " Video Indexing Based on Mosaic Representations ", David Sarnoff Research Center, Proceedings of IEEE, May 1998
 [2] Jiang Yu Zheng and Saburo Tsuji, "Generating Dynamic Projection Images for Scene Representatin and Understanding", Computer Vision and Image Understanding, Vol. 72, No. 3, December, pp.237-256, 1998
 [3] Heung-Yeung Shum and Richard Szeliski, "Panoramic Image Mosaic", Microsoft Research,

MSR-TR-97-23

[4] Dong-Gyu Sim and Rae-Hong park, "A Two-Stage Algorithm for Motion Discontinuity-Preserving Optical Flow Estimation", Computer Vision and Image Understanding, Vol. 65, No. 1, pp. 19-37, 1997
 [5] Richard Szeliski, Sing Bing Kang, "Direct Methods for Visual Scene Representation", June 24, 1995, IEEE Workshop
 [6] Steve Hsu and P.Anadan, "Hierarchical Representations for Mosaic-Based Video Compression", David Sarnoff Research Center
 [7] J.Kreyb, m.Roper, P.alshuth, Th.Hermes, and O.Herzog, "Video Retrieval by Still Image Analysis with ImageMiner", 1997, SPIE Vol.3022
 [8] Gulrukh Ahanger and Thomas D.C.Little, "A Survey of Technologies for Parsing and Indexing Digital Video", Journal of visual communication and image representation, Vol. 7, No.1, March, pp.28-43, 1996, ARTICLE NO.0004
 [9] Harpreet S.Sawhney and Serge Ayer, "Compact Representations of Videos Through Dominant and Multiple Motion Estimation", IEEE, April 11, 1996
 [10] Richard Szeliski, "Image Mosaicing for Tele-Reality Applications", Digital Equipment Corporation, Cambridge Research Lab, CRL 94/2, May, 1994
 [11] Richard Szeliski and James Coughlan, "Spline-Based Image Registration", Digital Equipment Corporation, Cambridge Research Lab, CRL 94/1, May, 1994
 [12] 최종원, 김진용, 유채곤, 서정수, 이성환, 황치정, "내용기반 비디오 파싱 기법을 통한 가변적 구성을 갖춘 브라우저의 설계 및 구현", 한국정보과학회 '97 봄 학술발표 논문집(B), Vol. 24, No.1, pp.19-22, 1997
 [13] 이동섭, 김재원, 배석찬, 이양원, "비디오 장면 변환 검출 알고리즘", 한국 퍼지 및 지능 시스템 학회 '98 춘계 학술대회 논문집, 제8권, 제1호, 1998.h.
 [14] 이동섭, 이양원, "뉴스 비디오의 장면분할을 위한 키 프레임 추출 기법", 한국정보과학회 춘계 학술발표 논문집, Vol.25, No.1, pp. 613-615, 1998