

## 효율적인 비디오 브라우징을 위한 파노라믹 영상 구축

신성윤\*, 박기홍\*, 표성배\*\*

# Construction of Panoramic Image for Efficient Video Browsing

Shin Seong-Yoon\*, Park Ki-Hong\*, Pyo Seong-Bae\*\*

### 요약

본 논문에서는 파노라믹 영상 구축을 위해 투시 투사에서 2차원 표면의 움직임을 더욱더 잘 표현하는 투시모델을 이용하여 구축하는 방안을 제시한다. 투시 모델에서 모션 파라미터는 현재 프레임의 영역과 이전 프레임에서 매칭된 영역 사이의 불일치 측정을 최소화함으로써 얻어진다. 프레임들은 고정된 참조나 시간에 따라 변하는 참조에 의해 동적 파노라믹 영상으로 구축된다. 실험을 통해서 본 논문에서 제시한 파노라믹 영상이 정지 영상이나 전체적인 비디오 보다 효과적으로 비디오의 내용들을 표현해 주고 크기 절감율도 높음을 알 수 있다.

### Abstract

In this paper, we present a construction method using perspective model which allows for the representation of the motion of a planar surface under perspective projection. The motion parameters are obtained by minimizing a disparity measure between the region in the current frame and the mapped region in the previous frame. The frames are constructed with dynamic panoramic image using a fixed reference or a time varying reference. By experiment, panoramic image proposed in this paper represents contents more efficiently than static image or video, and has a high volume reduction ratio.

▶ Keyword : 파노라믹 영상(Panoramic Image), 투시 모델(Perspective Model), 고정된 참조(Fixed Reference), 시간에 따라 변하는 참조(Time-Varying Reference)

• 제1저자 : 신성윤

• 접수일 : 2007.2.22, 심사일 : 2007.2.24, 심사완료일 : 2007. 3.19.

\* 군산대학교 컴퓨터정보학과 교수, \*\* 인덕대학 컴퓨터소프트웨어과 교수

※ 본 연구는 산업자원부와 한국산업기술평가단의 지역혁신인력양성사업으로 수행된 연구결과임

## I. 서론

최근 영상과 그래픽의 발전은 매우 폭넓게 진행되고 있다. 하드웨어와 소프트웨어의 지원으로 실영상과 더욱 유사한 영상을 볼 수 있게 되었고, 또한 사용자의 희망도 더욱 가세하게 되었다. 하지만, 멀티미디어의 막대한 자료의 처리 및 전송 그리고 정보의 이해가 실로 어려운 문제로 남아있다. 특히, 비디오는 많은 정보를 내포하고 있기 때문에, 사용자가 그 비디오의 내용물을 이해하기가 어렵다는 문제점을 가지고 있다. 이를 해결하기 위해 현재 많은 연구가 진행되어 오고 있으나, 효과적인 브라우징 방법은 개발되어 있지 않은 실정이다. 본 논문에서는 비디오의 내용물을 효과적으로 이해하기 위한 파노라믹 영상 생성 방법을 제시한다.

'파노라믹 영상'이라 함은 관련 있는 여러 영상들을 정합함으로써 하나의 새로운 영상으로 생성하는 것을 말하는데 흔히 '모자이크 영상' 이라고도 한다. 파노라믹 영상은 유사한 많은 영상(프레임)들을 하나의 영상으로 만들어 준다. 따라서 파노라믹 영상은 비디오의 대표 프레임에서 놓칠 수 있는 정보들의 모호성 문제점을 제거해준다. 하나의 장면을 대표하는 영상을 대표 프레임이라 하는데, 대표 프레임만을 가지고서 장면을 모두 이해하기에는 어려움이 많이 존재한다. 그렇지만, 파노라믹 영상은 이웃하는 영상을 모두 포괄하고 있게 되므로, 사용자는 파노라믹 영상을 통해서 한 장면을 총체적으로 이해할 수 있다. 파노라믹 영상 구축 과정의 기술들은 연속된 영상의 정렬, 중복, 통합의 요소들을 분석함으로써 이루어진다. 비디오 응용 기술로는 비디오 압축, 확장된 시각화, 비디오 향상을 내포하는 비디오 색인화, 탐색, 조정이 있다.

본 논문에서는 카메라를 통해서 입력 물을 받는다. 그러므로 카메라 파라미터의 인식을 위해서 투시모델을 이용하는데, 프레임 사이의 불일치 측정 방법을 제시하여 불일치를 최소화 하였다. 또한 파노라믹 영상을 생성하기 위하여 고정된 참조와 시간에 따라 변하는 참조를 각각 파노라믹 영상으로 제작하는 방법을 제시한다.

## II. 관련연구

파노라믹 영상 시스템을 구축하기 위한 논문은 커다란 이슈를 불러일으키고 있다. 현재 많은 연구가 진행되어 왔

으며, 더욱 발전하고 있는 실정이다. 많은 논문에서 차 영상, 영상의 좌표 변환 기법 및 영상 정렬과 움직임 측정에 대한 문제들을 제시하였다. 이 장에서는 파노라믹 영상 생성을 위한 기존의 방법들을 재조명한다.

Shum & Szeliski는 일반적인 평면에 투영하는 방법이 아닌 각 입력 영상을 변환 행렬과 관련하여 파노라믹 영상을 표현하는 것에 제안하였다[1]. 하지만 이 방법은 잡음에 매우 민감하다는 단점을 가지고 있다. Xiong과 Turkowski는 카메라를 90도로 회전하여 4개의 영상을 얻어 가상세계를 구축하는 방안을 제시하였다[2]. IBM사의 ImageMiner라는 시스템은 IBM사의 트레이드마크로서 비디오 분석과 정지 영상을 합한 시스템이다. 파노라믹 영상을 분석하는 단계로서 색, 질감, 윤곽 정보를 이용하여 객체를 인식하는 방법을 기술하였다[3]. 하지만 ImageMiner 시스템은 파노라믹 영상을 색, 질감, 윤곽 정보만을 가지고 객체를 인식하였기 때문에 동적 객체는 인식하지 못하는 문제점을 가지고 있다. 또 Michael Irani는 동적 파노라믹 영상에 대하여 제안하였지만, 움직이는 카메라와 동적 객체에 대해서 판별할 수 있는 방법은 기술하지 않았다[4]. 본 논문에서는 카메라와 동적 객체의 파라미터를 측정하고 이를 식별하는 알고리즘을 제안함으로써 정적 및 동적 파노라믹 영상을 구축하는 방법을 제안한다.

파노라믹 영상을 생성함에 있어 가장 쉬운 방법은 영상들 사이에서 단지 이동만 이루어졌다는 점이다. 이렇게 단지 이동되어진 영상들은 최소한의 제약으로 간단히 구현될 수 있고, 생성되어진 파노라믹 영상은 질이 높은 고해상도의 파노라믹 영상이 생성될 수 있으며, 계산 시간 또한 빠른 이점이 있다[5,6,7,8].

최근의 연구에서는 모자이크(파노라믹 영상화)으로 비디오 샷의 계층적 표현[9]과 칼만 필터링과 파노라믹 영상을 이용한 확률적 비디오의 안정화[10], 그리고 움직임 분석에 기반을 둔 효율적인 비디오 파노라믹 영상 구축[11] 등이 활발히 연구되고 있다.

## III. 파노라믹 영상 구축

카메라 움직임은 파라미터를 통하여 모델화 된다. 이를 위해, 어파인 모델과 투시 모델이 널리 사용된다. 어파인 모델은 다음 식 (1)과 같다.

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \dots\dots\dots (1)$$

그리고 투시 모델은 다음 식 (2)와 같다.

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{a_1x + a_2y + a_3}{a_7x + a_8y + 1} \\ \frac{a_4x + a_5y + a_6}{a_7x + a_8y + 1} \end{pmatrix} \dots\dots\dots (2)$$

식 (1)과 식 (2)에서 (x, y)<sup>T</sup>는 이전 프레임의 픽셀 위치이고 (u, v)<sup>T</sup>는 현재 프레임의 픽셀 위치이다. 어파인 모델은 정사영의 투사에서 2차원 표면의 움직임을 잘 표현하지만, 카메라의 거리에 관계가 적으므로 본 논문에서는 사용하지 않겠다. 오히려 파라미터 수가 많은 투시 모델을 본 논문에서 사용한다. 투시 모델을 투시 투사에서 2차원 표면의 움직임을 더욱더 잘 표현하기 때문이다.

모션 파라미터 a' = (a1, ..., an)은 현재 프레임의 영역 R과 이전 프레임에서 매칭된 영역 사이의 불일치 측정을 최소화함으로써 다음 식 (3)과 같은 방법으로 얻어진다.

$$\min_{a'} \sum_{r \in R} \|I(r, t) - I(T(r, a'), t - 1)\| \dots\dots\dots (3)$$

여기서 I(r, t)는 위치 r과 시간 t에서의 영상의 강도(명암도)를 표시하고, T(r, a')는 이전 프레임에 매칭된 위치를 나타낸다.

대표적인 파노라믹 영상 생성은 먼저 위의 투시모델을 이용한 카메라의 움직임을 계산한 후에 객체의 움직임을 인지하는 방법이다. 객체의 움직임을 위해서는 지역적 탐색이 이루어져야 한다. 지역적 탐색은 계산되어진 카메라의 움직임을 가지고서 두 영상간의 비교를 통해서 이루어지는데 계산식은 다음 식 (4)와 같다[12].

$$E(O) = \sum_{b \in W} [I_i(X + b) - I_j(X - u(k) + b + d_k)]^2 \dots\dots (4)$$

여기서 X는 한 픽셀의 공간적 좌표를 나타내고 Ii는 기준이 되는 영상을, Ij는 비교되어지는 영상의 순서를 나타낸다. b는 비교하는 한 윈도우 안에 픽셀이 몇 개인가를 이야기 하고, d는 비교되어지는 영상의 또 다른 부분 지역들을 나타낸다. 그리고 u(k)는 x축, y축으로의 방향벡터이다. 즉 카메라의 움직임의 거리만큼을 빼서 두 영상간의 에러 값을 계산한다.

본 논문에서는 파노라믹 영상을 생성하기 위해서 프레임 들은 2차원 좌표 시스템으로 할당된다. 좌표 시스템은 첫 번째 프레임이나 우선적인 프레임과 같은 고정된 참조나 현재 프레임 같은 시간에 따라 변하는 참조와 같다. 고정된 참조는 객체가 존재하지 않는 경우에 사용되고, 시간에 따라 변하는 참조의 경우에는 객체(사람)이 존재하는 경우로서 파노라믹 영상으로 구축된다. 객체의 존재 여부 판별은 하나의 새로운 주제이므로 본 논문에서는 제외하도록 하겠다.

고정된 참조는 객체가 존재하지 않는 경우로서 식 (5)를 사용하여 개별 프레임들이 파노라믹 영상으로 동적으로 통합된다.

$$P(r, t) = (1 - \alpha\beta(r))P(r, t - 1) + \alpha\beta(r)I(T(r, \sum a(t)), t) \dots\dots (5)$$

여기서 P(r, t)는 위치 r과 시간 t에서의 파노라믹 영상을 나타내고, Σa(t)는 참조 프레임과 현재 프레임 사이의 시간 간격 사이에 있는 연속적인 프레임들 사이의 전체 움직임 파라미터들의 합이다. T는 위치 r에서의 모션 파라미터 값을 나타내고 I는 시간 t와 모션파라미터 T에서의 영상을 나타낸다. α는 0과 1사이값과 같은 가중치를 나타내며, β(r)은 r이 배경이면 1이고 그렇지 않으면 0의 값을 갖는다.

현재 프레임에 대하여 객체가 존재하는 경우에는 시간에 따라 변하는 참조의 경우로서, 파노라믹 영상은 다음 식 (6)과 같이 구축된다.

$$P(r, t) = (1 - \alpha\beta(r))P(T(r, a'), t - 1) + \alpha\beta(r)I(r, t) \dots\dots (6)$$

하지만 객체를 제거한 배경 영상의 경우에는 식 (5)의 고정된 참조를 이용하여 배경 영상만 표현할 수 있다. 위와 같이 고정된 참조와 시간에 따라 변하는 참조로서 파노라믹 영상이 구축된다.

〈그림 3〉은 동적 객체가 존재하는 영상들로써 임의의 4개의 영상을 보여준 것이다.

#### IV. 실험

본 논문에서 제안한 파노라믹 영상 시스템을 구현하기 위해서 HITACHI Digital Zoom 카메라로부터 영상을 입력받았다. 실험에 사용되어진 입력 영상은 320\*240의 크기의 컬러 비디오 영상을 사용하였고, Pentium-IV에서 Visual C++6.0을 이용하였다.

〈그림 1〉은 모든 프레임들 중에서 임의의 4개의 영상만을 선택하여 나타내 준 것이다.

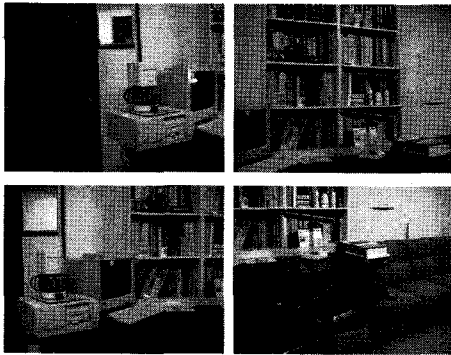


그림 1. 임의의 입력 영상 1  
Fig 1. Samples of Input Image 1

〈그림 2〉는 〈그림 1〉의 영상들을 입력받은 동영상 비디오로부터 고정된 참조에 의해 생성되어진 파노라믹 영상이다. 파노라믹 영상을 보면 상하좌,우가 상당히 길어진 것을 알 수 있으며, 이런 파노라믹 영상을 통해서 전체적인 비디오의 내용을 이해할 수 있다.



그림 2. 파노라믹 영상  
Fig 2. Panoramic Image

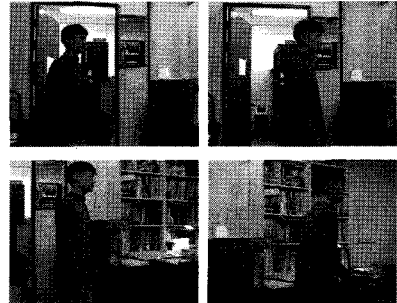


그림 3. 임의의 입력 영상 2  
Fig 3. Samples of Input Image 2

〈그림 4〉는 〈그림 3〉의 입력 영상에서 고정된 참조로서, 배경 영상만을 파노라믹 영상으로 구축한 것이고, 〈그림 5〉는 시간에 따라 변하는 참조로서, 동적 객체를 일정한 시간 별로 표현해 준 파노라믹 영상으로 구축되어진 영상이다.



그림 4. 배경의 파노라믹 영상  
Fig 4. Background Panoramic Image



그림 5. 전체적인 파노라믹 영상  
Fig 5. Wholly Panoramic Image

학교에서 캡처한 30초짜리 비디오 5개를 파노라믹 영상으로 구축한 결과는 〈표 1〉과 같다.

표 1. 파노라믹 영상 구축 결과  
Table 1. Results of Panoramic Image Construction

입력 비디오		파노라믹 영상		절감율
평균 프레임 수	평균 크기	평균 프레임 수	평균 크기	
900	8MB	97	750KB	90.8%
		89	725KB	91.14%
		92	735KB	91.03%
		94	737KB	91%
		88	712KB	92.31%
평균		92	731.8KB	91.07%

파노라믹 영상으로 구축한 결과 원래 비디오 크기인 8M(8192KB) 보다 평균적으로 91.07% 크기를 절감하면서 전체적인 내용을 표현함을 알 수 있다.

또한 지역적 탐색 방법과 제안된 방법의 절감율에 대하여 비교한 그래프는 <그림 6>과 같다.

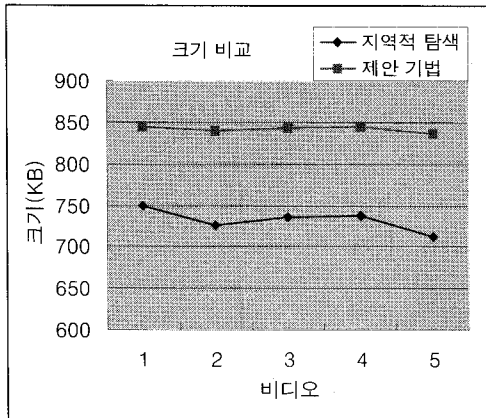


그림 6 절감율 비교  
Fig 6. Reduction Ratio Comparison

파노라믹 영상 구축 시 지역적 탐색 방법보다 본 논문에서 제시한 방법이 8M 비디오에 대해서 약 100KB 정도 저장 공간을 절감함을 알 수 있었다.

### V. 결론

본 논문에서는 파노라믹 영상 구축을 위해 이동하는 카메라를 통해 비디오 영상을 입력 받아 파노라믹 영상을 생성하는 방안을 제시하였다. 2차원 표면의 움직임을 더욱더

잘 표현하기 위하여 투시 모델을 이용하여 파노라믹 영상을 구축하였다. 객체가 존재하지 않는 고정된 참조와 객체가 존재하는 시간에 따라 변하는 참조의 경우에 파노라믹 영상으로 구축하는 방법을 제시하였다. 또한 전체적인 크기의 절감율도 본 논문에서 제시한 방법이 기존의 지역적 탐색 방법보다 훨씬 더 높음을 알 수 있었다. 많은 내용물을 포괄하고 있는 비디오를 브라우징 하는데 있어서, 파노라믹 영상이 대표 프레임보다 비디오의 한 장면을 더욱 효과적으로 표현할 수 있음을 알 수 있었다.

향후, 카메라의 변화가 심한 경우의 파노라믹 영상 생성과 파노라믹 영상으로 구축되어진 영상을 웹을 통하여 원 영상으로 재분할하는 방법에 대하여 연구할 것이며, 가상세계에 파노라믹 영상을 접목하여 학교 가이드 및 백화점 쇼핑몰 같은 응용분야에 대하여 연구 중이다.

### 참고문헌

- [1] Heung-Yeung Shum and Richard Szeliski, "Panoramic Image Mosaic", Microsoft Research, MSR-TR-97-23, 1997.
- [2] Y.Xiong, K.Turkowski, "Creating image-based VR using a selfcalibrating fisheye lens", Proc.CVPR' 97, 237-243.
- [3] J. Krey B, M.Roper, P.Alshuth, Th.Hermes, and O.Herzog, "Video Retrieval by Still Image Analysis with ImageMiner", SPIE Vol. 3022
- [4] M.Irani, P.Anandan, S.Hsu, "Mosaic based representation of video sequence and their applications", IEEE Proc ICCV'95, 605-611.
- [5] A. Tekalp, M.Ozkan, and M.Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration," in Proc. of the Int. Conf. on Acoust., Speech and Sigt. Proc., (San Francisco, CA), pp. III-169, IEEE, Mar. 23-26, 1992.
- [6] L. Teodosio and W.Bender, "Salient video stills : Content and context preserved," Proc. ACM Multimedia Conf., 1993.

- [7] 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.00004
- [8] Wonbo Yang, MoonSun Lim, Yangweon Rhee, "Video Mosaic System by Multi-Image", Korean Institute of Maritime Information & Communication Science, June, 1999, Vol.3 No.1, pp 104-108.
- [9] F. Odone, A. Fusiello, E. Trucco, "Layered Representation of a Video Shot with Mosaicing," Pattern Analysis & Applications, pp. 296-305, 2002
- [10] Andrew Litvin, Janusz Konrad, William C. Karl, "Probabilistic video stabilization using Kalman filtering and mosaicking," Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5022, pp. 663-674, 2003
- [11] T. MEI, X. S. HUA, H. Q. ZHOU, S. LI, H. J. ZHANG, "Efficient Video Mosaicing Based On Motion Analysis," ICIP 2005, September 11-14, Genoa, Italy. 2005.
- [12] Richard Szelisk, and H. Shum, "Creating full view panoramic image mosiacs and environment maps", In Proc. of SIGGRAPH, pages 251-258, 1997

**저자 소개**



**신성운**

2003년 2월 군산대학교  
컴퓨터과학과 이학박사  
2006년~현재 군산대학교  
컴퓨터정보과학과 전임강사  
<관심분야> 비디오 인덱싱, 비디오  
요약, 멀티미디어



**박기홍**

1995년 일본도쿠시마대학교 대학원  
정보시스템공학과 공학박사  
2001~현재 한국정보과학회  
호남·제주지부 지부장  
19856년~현재 군산대학교  
컴퓨터정보과학과 교수  
관심분야 : 정보검색, 센서네트워크



**표성배**

2004년 2월 숭실대학교  
컴퓨터학부 공학박사  
1992년~현재 인덕대학  
컴퓨터소프트웨어과 교수  
<관심분야> 멀티미디어, 비디오  
색인화, 비디오 요약