

# VideoGIS를 위한 동영상 자료구조화

## Video data Structuring for VideoGIS

김성삼\* · 김원만\*\* · 유환희\*\*\*

Kim, Seong Sam · Kim, Won Man · Yeu, Bock Mo · Yoo, Hwan Hee

### 요 지

최근 국내외에서 3D GIS 및 VideoGIS분야의 연구가 시작되고 있다. 동영상을 기반으로 하는 VideoGIS는 기존의 2D 및 2.5D GIS보다 더 현실감있고 생동감있는 정보의 제공이 가능할 것으로 기대된다. 그러나 동영상의 경우, 그 양이 방대할 뿐 아니라, 본질적으로 순차적인 특징을 가지고 있어 자료 사용(브라우징 및 검색)에 있어 사용자가 필요한 정보에 접근하기 위해서는 반복적인 검색기능을 수행해야 하는 문제점이 있다. 동영상 자료의 접근성 개선 및 VideoGIS의 기초 자료로 활용하기 위해 본 논문에서는 동영상 자료구조화에 관한 연구를 수행하였다.

### 1. 서론

빠르게 발전하는 컴퓨터 관련 기술과 더 넓어진 네트워크 밴드 폭을 통하여 사용자들은 이미지, 오디오, 특히 비디오를 포함하는 멀티미디어 환경에서 더 현실성 있는 정보를 검색하기를 요구하고 있다. GIS도 이러한 변화 속에 대응하여 최근 VideoGIS라는 새로운 기술이 등장하게 되었다. 가상현실에 공간정보를 부여하여 또 다른 현실을 창출해내는 3차원 GIS 기술과 더불어 동영상을 기반으로 하는 VideoGIS는 보다 현실성 있는 정보를 제공하는 GIS의 분야로 발전할 가능성이 높다. 그러나 국내외에서는 아직 그래픽, 동영상, 소리, 음향 등을 편집하고 수정하는 멀티미디어 개념차원에서 연구가 많이 진행되고 있으나, GIS 개념의 도입은 아직 미약한 상태이다. 따라서 본 논문에서는 VideoGIS를 구현하는데 필요한 비디오

동영상자료에 대한 기초 개념과 자료 구조화 및 활용에 대해 간단한 시스템 구현을 통해서 제시하고자 한다.

### 2. VideoGIS 개념

사람들은 방송이나 영화 그리고 비디오 테이프, 인터넷과 같은 매체를 통하여 정보를 제공받고 있으며, 이때 우리가 제공받는 영상정보를 비디오라고 한다. 이러한 비디오 정보에 GIS기능을 더하여 우리가 원하는 정보를 추출하고, 필요로 하는 사용자들에게 정보를 제공하고자 한다. 그러나 지금까지의 GIS는 주로 공간적인 위치와 관련되어 있는 정보들을 가시적으로 관리하고, 분석하며, 사용자에게 2차원적으로 보여주고 관리할 수 있도록 하고 있다. 그러나 VideoGIS는 기존의 컴퓨터 그래픽기술을 기반으로 하는 GIS와는 달리 비디오 동영상을 기반으로 공간자료를 분

\*경상대학교 건설공학부 도시공학과 박사과정 · 055-751-5321(E-mail : kimss333@netian.com)

\*\*경상대학교 건설공학부 도시공학과 석사과정 · 055-751-5321(E-mail : ok016@hotmail.com)

\*\*\*경상대학교 건설공학부 도시공학과 교수 · 공학박사 · 055-751-7321(E-mail : hhyoo@nongae.gsnu.ac.kr)

석하고 의사결정을 도와주는 시스템이다. VideoGIS를 통하여 사용자들은 기존의 GIS로부터 얻을 수 있었던 정보 뿐 아니라 공간 지형 자료의 현실성 있는 정보를 확인할 수 있다. 이러한 영상 정보를 통해서 사용자는 현실 세계와 컴퓨터상으로 표현되는 지형 공간 데이터의 관계성을 더 정확하고 빠르게 알아볼 수 있다(유재준 등, 2002).

### 2.1 비디오 자료 구조

비디오 자료는 히스토그램 분석을 통하여 계층적 트리 형태의 자료로 구조화될 수 있다(그림 1). 비디오 샷(shot) 계층은 카메라로부터 기록되어지는 프레임의 연속 장면으로 비디오를 블록화 한다. 비디오 씬(scene) 계층은 동질한 정보나 자료를 갖는 비디오 샷의 집합으로 정의 되고, 시간적으로도 연속 된다. 비디오 그룹 계층은 샷과 씬 사이의 중간 계층으로 비디오 그룹에서 샷은 시각적으로는 유사하지만 시간적으로는 일치하진 않는다. 마지막으로 비디오 계층은 최상위 계층으로서 비디오 자료의 모든 요소를 포함하고 있다(Rui 등, 1999).

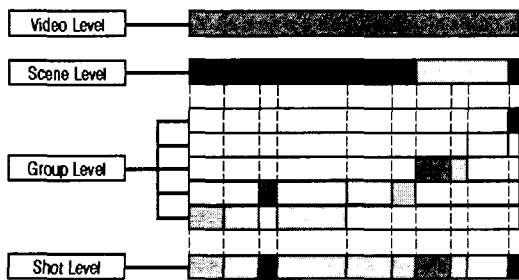


그림 1. 비디오 자료 계층적 트리 구조

### 2.2 비디오 분할 처리 과정

비디오 자료를 구조화하기 위한 첫 단계는 분할 처리(segment process) 과정이다. 대부분의 분할 처리는 연속되지 않는 장면에서 비디오를 분할하고 있다. 비디오 분할 방법의 대부분은 이미지 기반(image-based), 오디오기반(audio-based),

그리고 텍스트 기반(text-based)으로 분류된다(Li 등, 1996 ; Lienhart, 2000). 다른 분할 방법들도 세 가지 범주의 유형을 혼합시킨 혼합형이다. 오디오 기반의 방법은 오디오 에너지 계층(audio energy level)의 변화 중 갑작스럽게 변화하는 것을 임계값으로 설정하여 비디오를 분할하며, 텍스트 기반 방법은 표제 텍스트의 변화를 찾아서 비디오를 분할하고, 현재는 이미지 기반 방법이 가장 많이 이용되고 있는데, 이 방법은 이미지에 대한 히스토그램분석을 통해 샷 경계를 검출한다.

비디오 자료를 구조화하기 위해서 처음 비디오 특성으로 컬러 히스토그램을 추출한다. 추출된 컬러 특성을 기초로 프레임 간 차이 값을 구하고, 샷 경계를 찾는다(그림 2).

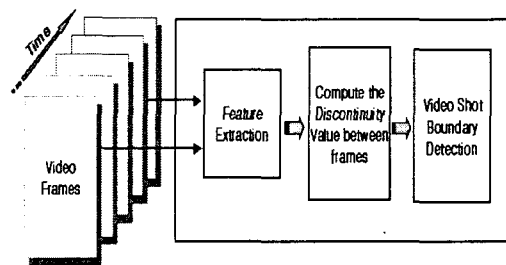


그림 2. 비디오 분할 처리 과정

### 2.3 비디오 샷(shot) 경계 검출

비디오 샷 경계(shot boundary)는 연속적인 비디오 프레임  $f_x$ 와  $f_{x+1}$ 사이의 프레임 간 차이가 임계값 보다 더 클 경우에 설정된다(Yu 등, 2001 ; Browne 등, 2000).

$$z_i(f_x, f_y) = \sum_{j=1}^N |Hist_{x,i}(j) - Hist_{y,i}(j)|, \quad (1)$$

$$z(f_x, f_y) = \sum_i (z_i(f_x, f_y) \times w_i)$$

식(1)에 의해서 프레임 간 차이를 계산하고, 그 결과 값과 식(2)에서 구해진 임계값을 비교하여 샷의 경계를 검출한다.

$$T = t_{opt} \times \frac{z_{max}}{q}$$

$$z(f_x, f_{x+1}) \begin{cases} > T \rightarrow \text{Shot boundary occurs} \\ \leq T \rightarrow \text{Not a shot boundary} \end{cases} \quad (2)$$

## 2.4 비디오 그룹(group) 구조화

비디오 그룹은 비디오 샷의 집합으로 정의된다. 비디오 그룹과정에서 임계 값으로  $T_{keyframe}$ 과  $T_{temporal}$ 을 필요로 한다. 각 샷의 첫 번째 프레임이  $T_{keyframe}$ 으로 정의되고,  $T_{temporal}$ 은 식(3)으로 계산되어 진다.

$$T_{temporal} = \frac{\sum_{i=1}^m length(s_i)}{m} \times K, \quad (3)$$

$$s_i = [f_x, f_y], \quad length(s_i) = y - x$$

샷과 그룹사이의 차이가  $T_{keyframe}$ 보다 작을 경우, 그리고 현재 샷과 그룹의 마지막 요소사이의 시간간격이  $T_{temporal}$ 보다 작을 경우에 같은 그룹으로 분류한다. 그렇지 않으면 현재 샷에 대하여 새로운 그룹을 만들어 분류한다. 샷과 그룹의 차이는 식(4)으로 계산된다.

$$\text{shot-to-group distance,}$$

$$z_g(s_x, g_i) = \frac{z(k_x, k_y) + z(k_x, k_z)}{2} \quad (4)$$

video shot :  $s_x$ ,

video group :  $g_i = (s_y, s_u, \dots, s_z)$

## 2.5 비디오 씬(scene) 구조화

비디오 씬 역시 하위 구조인 비디오 그룹의 집합으로 정의 된다. 그룹을 씬으로 구조화하기 위해서는 두 가지 선행 작업이 필요하다. 첫째, 가장 처음의 비디오 샷을 가지고 비디오 그룹을 시간적 순서에 따라 정렬해야 한다. 둘째, 각 씬의 시간 간격과 그룹에서 처음과 마지막 요소의 시간 간격을 비교한다. 그룹의 정렬 후

세 가지 경우에 대하여 비디오 씬(scene)으로 할당된다.

Case (1), 그룹의 시간 간격이 씬의 시간 간격에 완전히 겹쳐지는 경우 같은 씬으로 간주한다.

Case (2), 그룹의 시간 간격이 일부 씬의 시간 간격에 겹치는 경우에 같은 씬으로 분류된다. 그러나 이 경우 씬의 시간간격을 조정해야 한다.

Case (3), 그룹의 시간 간격이 씬의 시간 간격에 겹쳐지지 않는 경우 새로운 씬을 생성한다.

## 3. VideoGIS 시스템 구현

### 3.1 동영상 Play

VideoGIS는 동영상을 기반으로 사용자에게 정보를 제공해주는 시스템이다. 과거 주로 문자와 간단한 그림으로 정보를 알려 주었다. 현재 문자나 그림만으로는 정보 전달의 한계가 있으므로 기존의 문자 정보와 동영상 play 기능을 혼합하여 정보를 제공할 경우 더욱 현실감 있는 정보제공이 가능하다(그림 4).

### 3.2 VideoGIS를 위한 자료구조화

동영상은 시각적 정보만 제공하므로 GIS 기초 자료로 사용하기 위해서는 동영상을 분할하고, 편집하는 새로운 자료 생성과정이 필요하다. GIS 기초 자료를 얻기 위해서 동영상을 프레임별로 저장하여 정지영상으로 만들고, 동영상의 shot 정보 및 keyframe 정보를 추출하여 동영상을 유사한 종류로 편집한다. 이 편집된 자료에 구조화과정과 속성을 추가해주면 VideoGIS에 필요한 자료 구조가 구축된다. 동영상은 본질적으로 순차적인 특징을 가지고 있어 필요한 부분에 접근하기 위해서는 고속검색(fast forward or fast backward)을 반복해야 한다. 그러나 동영상을 구조화하면 시간적, 편리성, 접근성 측면이 개선된다.

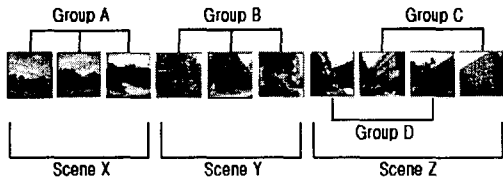


그림 3. 비디오 그룹 및 씬 구조화

### 3.3 이미지 정보 검색

동영상에서 문제지역의 화면이 정지했을 때 맵 창(map window)에 정지영상을 출력하고 정지영상에서 문제 시설물을 벡터로 표시하고, 표시된 시설물의 속성을 입력할 수 있는 기능이다. 사용자들은 여러 동영상에서 문제 시설물이 있는 이미지만 캡처하여 일괄적으로 저장해놓고, 유사 시설물을 검색하여 시설물에 대한 필요한 정보를 처리하는 기능이다. 도시 시설물을 관리할 경우 이 기능을 통해서 자신이 관리하고자하는 지역이나 시설물에 대하여 직접 방문하기 전에 일차적으로 위치나 간단한 시설물 정보의 취득이 가능할 것으로 기대된다.

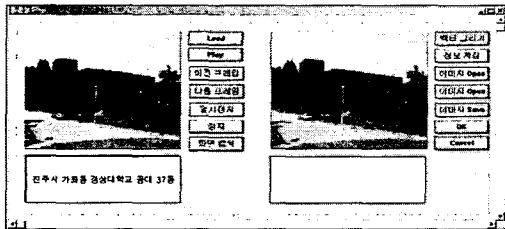


그림 4. 동영상 처리 인터페이스

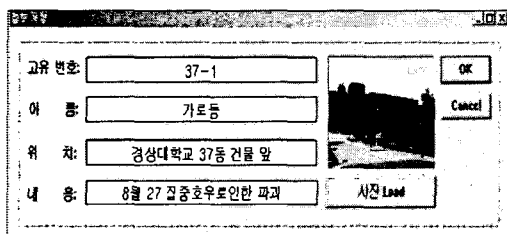


그림 5. 정보 입력

## 4. 결론

GIS분야에 비디오 동영상을 이용하기 위한 노력의 일환으로 VideoGIS에 대

한 개념과 구조화 방법 등에 대해 연구한 결과 다음과 같은 결론을 얻었다.

현실세계에서 이뤄지고 있는 다양한 실세계 정보를 그대로 저장하여 사용자에게 전달하여 줄 수 있는 비디오 동영상은 기존의 GIS자료와는 다른 새로운 차원에서 활용될 수 있는 가능성을 많이 갖고 있으며 이러한 비디오 동영상은 단순히 play하여 현황을 파악하는 단계를 넘어서 자료를 구조화하고 속성을 연계시킴으로서 비디오와 속성이 통합된 새로운 차원의 VideoGIS를 구현할 수 있는 가능성을 찾을 수 있었다.

## 참고 문헌

1. 유재준, 최경호, 장병태, 이종훈(2002) Video GIS 기술 동향, ETRI.
2. P. Browne, A. F. Smeaton, N. Murphy, N. O'Connor, S. Marlow, C. Bernt(2000) Evaluating and Combining Digital Video Shot Boundary Detection Algorithms, *In Proceedings of the Irish Machine Vision and Image Processing Conference (IMVIP)*.
3. W. Li, S. Gauch, J. Gauch, K. M. P. Pua(1996) VISION: A Digital Video Library. *In ACM Digital Libraries*, pp. 19-27.
4. R. Lienhart(2000) Automatic Text Segmentation and Text Recognition for Video Indexing, *In Multimedia Systems Hrsg : ACM Berlin-Heidelberg : Springer-Verlag*, S. 69-81, volume.
5. Y. Rui, T. S. Hunag, and S. Mehrotra(1999) Constructing Table-of-Content for Video. *In ACM Multimedia Systems Journal, Special Issue Multimedia Systems on Video Libraries*, volume 7, no. 5, pp. 359-368.
6. J. Yu, M. D. Srinath(2001) An efficient method for scene cut detection, *In Elsevier Pattern Recognition Letters*, volume 22, pp. 1379-1391.