

## ART2를 이용한 장면 전환 검출

임혁순<sup>\*</sup> 박상성 문호석 이만희 장동식  
고려대학교 산업시스템정보공학과  
{diagurtns<sup>\*</sup>, hanyul, bawooi, manhee, jang}@korea.ac.kr<sup>+</sup>

### Scene Change Detection using ART2

Hyuksoon Im<sup>\*</sup> Sangsung Park Hoseok Moon Manhee Lee Dongsik Jang  
Dept. of Industrial Systems and Information Engineering ,Korea University

#### 요 약

디지털 비디오에 있어서 멀티미디어 데이터베이스 및 검색 시스템 구축을 하기 위해서는 비디오를 여러 개의 장면으로 분할하는 기술이 요구된다. 본 논문에서는 다양한 종류의 장면 전환을 검출하기 위해 기존의 규칙기반이 아닌 신경망 이론을 접목하여 자율학습과 실수값 입력이 가능한 ART2를 이용하였다. 매 프레임마다 발생할 수 있는 변동의 폭을 줄이기 위해 MPEG 동영상의 DC에 해당하는 값만을 이용하고, 프레임마다 색상의 분산값을 이용하여 Plateaus 구간을 검출한 다음 Plateaus 구간에 해당하는 프레임들에 대해서만 프레임차이, 히스토그램차이, 상관계수 등의 특징치를 추출하여 ART2에 특징벡터를 입력하여 장면 전환을 검출하였다.

#### 1. 서 론

최근 디지털 비디오와 같은 영상 데이터에서 중요한 정보를 추출하고 효율적으로 검색을 하기 위해서 지원해야 할 기능은 비디오의 내용을 인덱싱하여 사용자가 원하는 정보에 비순차적으로 접근할 수 있게 하여야 하며 이를 위해서는 한 편의 비디오에서 장면이 바뀌는 시점 단위로 분할하고 대표 프레임을 추출하기 위한 장면 전환 검출 기술의 필요성이 중요시 된다. 즉, 디지털 비디오에 있어서 멀티미디어 데이터베이스 및 검색 시스템 구축을 위한 첫 번째 단계로써 비디오를 여러 개의 장면(Scene or Shot)으로 분할하는 기술이 요구된다. 기존 연구들에서 대부분 공통적으로 안고 있는 문제점은 장면 전환을 선연하기 위한 임계치 설정에 관한 문제이다. 임계치 설정에 따라 상이한 결과가 도출되기 때문에 실제 기존 연구에서 제시된 방법을 적용할 때 어려움이 발생한다. 또한 실제로 장면 전환은 아니지만 장면 전환과 유사한 특징치를 가지는 프레임들을 구별하기 위해 제시한 방법들이 비디오의 편집 과정에서 발생할 수 있는 점진적 장면 전환을 검출하는데 좋지 못한 결과를 보이고 있다. Nagasaka 와 Danaka가 제시한 방법은 확대나 이동과 같은 카메라 효과에는 비교적 강인하나 페이딩과 같은 편집 효과에는 취약한 결과를 보인다[1]. Zhang 등이 제시한 방법은 상이한 프레임 사이에 칼라 분포가 유사하거나 플래시등에 의해 급격한 변화를 보일 때 검출에 실패하는 경향이 많다[2]. 본 연구에서는 다양한 종류의 장면 전환을 검출하기 위해 기존의 규칙기반이 아닌 신경망 이론을 접목하여 자율학습과 실수값 입력이 가능한 ART2를 이용하였다. 매 프레임마다 발생할 수 있는 변동의 폭을 줄이기 위해 MPEG 동영상의 DC에 해당하는 값만을 이용하고, 프레

임마다 색상의 분산값을 이용하여 Plateaus 구간을 검출한 다음 Plateaus 구간에 해당하는 프레임들에 대해서만 프레임차이, 히스토그램차이, 상관계수 등의 특징치를 추출하였다. 이렇게 추출된 특징벡터를 ART2에 입력하여 장면 전환과 비전환에 해당하는 두 개의 출력 클러스터의 소속 프레임을 이용하여 장면 전환을 검출하였다.

#### 2. 특징벡터 추출

2.1 최소 분산 영역 그래프  
DC 영상에서 분산을 계산하여 시퀀스로 나타내면 디졸브 구간에서 이론적으로는 "U"자 형태를 나타내야 하지만 실질적으로는 그렇지 않은 경우도 있고 또한 디졸브가 아닌 구간에서도 "U"자 형태가 나타나는 경우도 있다. 동영상내에 디졸브가 하나도 없는 경우에도 "U"자 형태를 나타내는 구간이 존재한다. 본 연구에서는 기본적으로 분산을 이용하되 디졸브를 검출하는 것이 아니라 디졸브의 가능성이 있거나 디졸브와 유사한 형태를 나타내는 구간을 제거하여 분산 시퀀스 내에서 Plateaus 형태를 갖는 부분과 지역 최대치를 갖는 부분만을 얻어낸다. 분산 그래프에서 지역 최대치와 Plateaus 특성을 갖는 부분을 강조하고 디졸브가 비교적 짧은 구간에 발생하는 구간에 디졸브 특성을 강조하기 위하여 DC 영상을 그림 2.1과 같이 서로 중첩된 5개의 영역으로 구분하여 각 영역별로 분산값을 계산한 다음, 최소 분산값을 가는 영역의 분산값을 이용해 시퀀스를 생성한다. 전체 분산 그래프와 최소 분산 영역 그래프를 비교해보면 전체 분산그래프에서는 디졸브 특성인 "U"자 형태가 거의 나타나지 않지만 최소분산 영역 그래프에서는 선명하게 잘 나타난다. 또한 전체 분산 그래프에서는 디졸브 구간이 뚜렷하게 나타나지만 최소 분산 영역 그래프에서는 잘 나타나지 않는다. 그러므로 이 두 가지 그래프를 모두

이용하여 Plateaus 구간과 지역 최대치 구간을 구하는 것이 좀 더 좋은 결과를 얻을 수 있다.

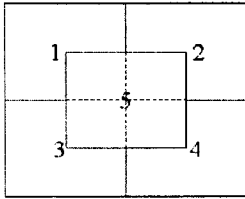


그림 2.1 영역 분할

2.2 Plateaus 구간 검출

ART2에 입력될 프레임을 결정하기 위해 앞장에서 구한 최소 분산 영역 그래프에서 Plateaus 와 지역 최대값을 갖는 위치를 구해야 한다. Plateaus 는 분산값의 변화가 거의 없는 평탄한 구간이고 지역 최대값은 "U"자 형태와는 반대되는 모양을 가진 구간의 극점이 된다. 먼저 그래프의 작은 변동을 없애기 위하여 크기가 m인 평활화 필터를 적용하여 식 2.1과 같은 새로운 시퀀스  $A_i$  를 만든다:

$$A(i) = \frac{1}{m} \sum_{j=i-\frac{m}{2}}^{i+\frac{m}{2}} V(j) \tag{2.1}$$

where, m : size of averaging filter

$V(j)$  : sequence of variance

$$i = \left\lfloor \frac{m}{2} \right\rfloor, \left\lfloor \frac{m}{2} \right\rfloor + 1, \left\lfloor \frac{m}{2} \right\rfloor + 2, \dots, N-1-\left\lfloor \frac{m}{2} \right\rfloor$$

$$j = 0, 1, 2, \dots, N-1$$

N = length of sequence

그 다음, 식 2.2과 같이 이 시퀀스에 일차미분을 수행하여 미분값이 임계치 이하인 부분을 일차적으로 Plateaus 구간으로 간주하는 시퀀스  $p$ 를 생성한다.

$$p(i) = \begin{cases} 0 & \text{if } D(i) < T_d \\ A(i) & \text{otherwise} \end{cases} \tag{2.2}$$

where,  $D(i) = |A(i) - A(i+1)|$ ,  
 $T_d$ : threshold value

이 구간에는 지역 최대, 지역 최소, 그리고 상승곡면이나 하강곡면에서 기울기가 완만한 구간이 모두 포함된다.

상세한 알고리즘은 다음과 같다.

- step1. 시퀀스의 값을 순차적으로 조사한다.
- step2. 시퀀스의 값이 Plateaus 에 속하지 않으면 step1로 넘어간다. 시퀀스의 값이 Plateaus 에 속하면 현재의 위치에서부터 Plateaus구간의 끝이 나올 때까지 진행한 다음 구간의 폭을 계산한다.

- step3. Plateaus 구간의 폭이 임계치보다 크면 step1로 넘어간다. Plateaus 구간의 폭이 임계치보다 작으면 step4로 진행한다.
- step4. Plateaus 구간의 처음에서 앞으로 나아가면서 비 Plateaus구간의 처음 위치를 구한다.
- step5. Plateaus 구간의 끝에서 뒤로 나아가면서 비 Plateaus구간의 끝 위치를 구한다.
- step6. Plateaus 구간내에서 분산의 최대값을 계산한 다음, 이 값이 비Plateaus 구간의 처음위치에서의 분산값과 비Plateaus구간의 끝위치에서의 분산값보다 모두 작은 경우에 이 구간이 지역 최대임을 의미한다. 그렇지 않은 나머지 경우는 모두 비 Plateaus구간으로 변경한다.

3. ART2 학습 알고리즘

3.1 ART2의 구조

ART 모델은 주의 서브시스템(attentional subsystem)과 적응 서브시스템(orienting subsystem)으로 되어 있고, 주의 서브시스템은 입력패턴을 받아들여 비교하는 F1층과 인식된 결과를 출력하는 부분인 F2층으로 되어 있다. F1과 F2층은 각 노드에 활성화된 패턴을 저장하기 때문에 STM(Short Term Memory)이라 하고, F1과 F2사이의 상황, 하향 연결가중치를 LTM(Long Term Memory)으로 나타낸다. ART 신경회로망은 LTM의 연결 가중치를 조절함으로써 새로운 패턴에 대한 학습을 수행한다. 주의 서브시스템은 F1층에 입력된 패턴이 학습된 F2층과의 매칭에 실패했을 때 활성화되어 F2층에 차단 신호(reset signal)를 보내어 활성화를 억제하여 적당한 매칭이 발견되거나 새로운 카테고리가 생성될 때까지 나머지 카테고리를 차례로 검사하도록 한다. 전형적인 ART2의 구조는 그림 3.1과 같다. F1층은 여섯가지 형태의 유닛(W, X, U, V, P, Q)으로 구성되어 있으며 각각의 유닛은 입력 패턴의 크기와 같은 n 개의 벡터로 이루어져 있다 [3][4][5]. ART2의 F1 층은 리셋 매커니즘에 필요한 상황과 하향 신호의 비교와 더불어 정규화와 잡음억제의 조합을 포함한다. F2층은 클러스터 유닛으로 경쟁학습이 이루어지는 층으로 F1층의  $P_i$  유닛과  $Y_j$  유닛의 연결강도가 상황 연결 강도와 하향 연결 강도로 연결되어 있다. 경쟁학습에서는 winner-take-all 방식이 적용되어 오직 한 개의 유닛만이 주어진 입력벡터에 대해서 학습될 권한을 갖게 된다.

ART2의 경쟁층은 여러 가지 셋으로 구성된 다층구조로 이루어져 있으며 추가로 만들어진 노드들은 다음과 같은 역할을 수행한다. 첫째로 잡음 억제를 허용하고, 둘째로 패턴벡터에서 중요한 부분을 향상시킬 수 있는 정규화 과정이 포함되고, 마지막으로 리셋 매커니즘에 필요한 상황과 하향 신호의 비교 기능이다.

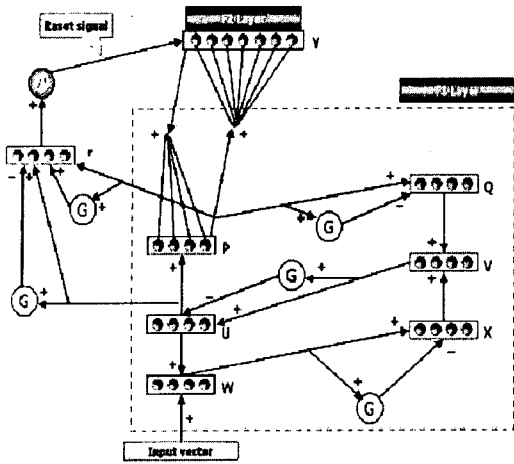


그림 3.1 ART2 구조

3.2 ART2의 학습 알고리즘

ART2의 입력부터 출력까지 이르는 과정은 다음과 같다.

Step 0. 파라미터의 초기화

$$a, b, c, d, e, \alpha, \beta, \theta$$

Step 1. Do Step 2-12 N\_EP time

(Perform the specified number of epochs of training.)

Step 2. For each input vector s, do steps 3-11

Step 3. Update F1 unit activations:

$$u_i = 0, x_i = \frac{s_i}{e + ||s||}$$

$$w_i = s_i, q_i = 0, p_i = 0, v_i = f(x_i)$$

Update F1 unit activations again

$$u_i = \frac{v_i}{e + ||u||}, w_i = s_i + \alpha u_i, p_i = u_i,$$

$$x_i = \frac{w_i}{e + ||w||}, q_i = \frac{p_i}{e + ||p||},$$

$$v_i = f(x_i) + b f(q_i).$$

Step 4. Compute signals to F2 units :

$$Y_j = \sum b_{ij} p_i$$

Step 5. While reset is true, do Step 6-7

Step 6. Find F2 unit Yj with largest signal.

(Define J such that Yj > Yk for k=1, ..., m)

Step 7. Check for reset :

$$u_i = \frac{v_i}{e + ||u||}, p_i = u_i + \alpha p_i, r_i = \frac{v_i + \beta p_i}{e + ||u|| + c ||p||}$$

if || r || < p - e, then yj = -1 (inhibit J)

(reset is true ; repeat Step 5);

if || r || > p - e, then

$$w_i = s_i + \alpha u_i, x_i = \frac{w_i}{e + ||w||},$$

$$q_i = \frac{p_i}{e + ||p||}, v_i = f(x_i) + b f(q_i).$$

Reset is false ; proceed to Step 8.

Step 8. Do Step 9-11 N\_IT times.

(Perform the specified number of learning iterations)

Step 9. Update weights for winning unit J :

$$t_{ji} = \alpha u_i + 1 + \alpha \beta (d - 1) t_{ji}$$

$$b_{ij} = \alpha u_i + 1 + \alpha \beta (d - 1) b_{ij}$$

Step 10. Update F1 activations :

$$u_i = \frac{v_i}{e + ||u||}, w_i = s_i + \alpha u_i, p_i = u_i,$$

$$x_i = \frac{w_i}{e + ||w||}, q_i = \frac{p_i}{e + ||p||}, v_i = f(x_i) + b f(q_i).$$

Step 11. Test Stopping condition for weight updates.

Step 12. Test Stopping condition for weight epochs.

4. 구현

현재 개발 중인 이 시스템은 ART2를 이용하여 장면 전환을 하는 시스템이다. 실험과 결과분석을 위하여 Window 2000, Pentium IV, Ram 1G의 실험 환경을 설정하여 VC++ 툴로 개발 중이다.

5. 결론

본 논문에서는 ART2를 이용하여 비디오의 장면 전환을 검출하는 알고리즘이 제시되었다. 비디오의 장면 전환 검출 기술은 대용량의 멀티미디어 정보를 재구성하여 효율적인 검색 및 데이터베이스 관리를 위한 첫 번째로 해결해야 할 과제이다. 본 논문에서는 다양한 종류의 장면 전환을 검출하기 위해 신경망 이론을 적용하여 자율학습과 실시간 입력이 가능한 ART2를 이용하였다.

향후 과제로는 오브젝트의 이동을 감지하여 움직임의 보상할 수 있도록 하여 장면 전환으로 잘못 검출되는 오류를 줄이고 다양한 비디오 이펙트에 대응하기 위한 디졸브의 이론적 모델의 재정립이 요구된다.

참고문헌

[1] A. Nagasaka and Y. Tanaka, "Automatic video indexing and full-video search for object appearances," in Visual Database Systems II, pp. 113-127, 1992.  
 [2] H. J. Zhang, A. Kankanhalli, S.W.Smoliar, "Automatic partitioning of Video," ACM/Springer Multimedia Systems, vol. 1, No. 1, pp. 10-28, 1993.  
 [3] Laurene Fausett, (1994), Fundamentals of Neural Networks, Prentice Hall.  
 [4] James A. Freeman, David M. Skapura, (1991), Neural Networks Algorithms, Applications, and Programming Techniques, Addison-Wesley.  
 [5] Abhihit S. Pandya, Robert B. Macy, (1995), Pattern Recognition with Neural Networks in C++, CRC Press.