시트콤 동영상에서 MPEG-7 시각 기술자를 이용한 Scene 배경의 자동 분류 방법

전재욱, 손대온°, 낭종호
서강대학교 공과대학 컴퓨터학과
{elligio, maxson°}@mlneptune.sogang.ac.kr, jhnang@ccs.sogang.ac.kr

An Automatic Scene Background Classification Scheme for Sitcom Videos Using MPEG-7 Visual

Jaewook Jeon, Daeon Shon°, Jongho Nang
Department of Computer Science, Sogang University

요 약

시트콤 동영상은 고정된 배경을 갖는 줌 아웃에 연이어 오는 줌 인으로 구성되어 있고, 또한 촬영되는 배경의 수는 한정되어 있는 특성이 때문에, 이러한 배경의 시각적 특성을 사용하여 배경들을 학습시키고 자동으로 분리시킬 수 있다. 본 논문에서는 신경망의 일종인 LVQ[1]를 사용하여 이러한 종류의 비디오 동영상에 대한 자동 배경 분류 방법을 제안한다. 우선, MPEG-7 시각 기술자를 이용하여 신(scene) 배경의 시각적인 특성을 추출하고 이러한 시각적 특성을 미리 제작자에 의해서 주어진 배경 정보로서 LVQ를 학습시킨다. 학습이 진행되면서 특정 배경의 시각적 특성은 LVQ의 가중치로서 표현되며, 다른 배경을 자동으로 분류하는데 사용된다. 제안된 LVQ기반의 분류 방법을 사용한 두 종류의 시트콤 동영상에 대한 실험 결과는 분류에 대한 어떠한 하드코딩 없이 80~90%의 정확도로 시트콤 동영상의 배경을 자동으로 분류한다.

1. 서 론

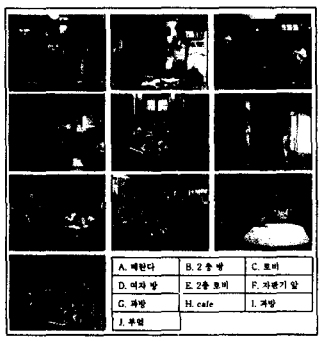
컴퓨터의 급격한 발달, 네트워크의 속도의 증가 그리고 저장 장치의 대용량화로 멀티미디어정보의 사용이 증가되고 있다. 동영상 데이터에 대한 서비스를 실시하기 위해서는 대용량의 동영상 데이터를 인덱싱, 저장, 그리고 부가적인 정보를 입력하여 사용자들이 원하는 항목을 쉽게 찾을 수 있도록 해야 한다. 이렇게 인덱싱된 동영상 데이터에 대한 고급 수준의 부가적인 정보를 자동으로 추출하는 동영상 분석 방법이 필요하다. 그러나 저급 수준 정보를 통계적인 방법을 통하여 사용할 경우, 다양한 동영상의 데이터에서 사용자가 원하는 고급 수준의 정보를 찾아내는 것에는 한계가 있다.

본 논문에서는 시리즈형식의 시트콤 동영상에 대하여 신 단위의 배경을 분류하기 위해서 시각적인 특성을 추출하고, 이를 바탕으로 이미 분류된 신들을 바탕으로 학습을 통해 신 단위의 배경을 자동으로 분류하는 방법을 제안한다. 배경을 자동으로 분류하기 위해서 시트콤 동영상의 배경의 특성을 분석하였다. 시트콤 동영상은 정형화된 배경[2]을 지니고 있으며, 이러한 정형화된 배경은 줌 아웃과 줌 인을 반복하는 특징을 지니고 있다. 이러한 특징을 이용하기 위하여 줌 아웃에서 배경을 대표하는 시각적 특성들을 추출한다. 추출한 시각적 특성을 이용하여 LVQ라는 신경망을 학습시킨다. 신경망은 배경에서 추출한 MPEG-7 시각 기술자들의 정보를 입력과 가중치로 갖게 된다. 그 결과 하드 코딩된 프로그램이 아닌 학습된 신경망에 의한 것이므로, 처음에는 하드 코딩된 프로그램보다 성능이 떨어진다. 그러나 점차로 학습이 되면서, 시트콤 동영상의 종류에 상관없이 배경에 대한 분류성능이 높아짐을 실험을 통하여 확인하였다.

• 본 연구는 한국과학재단에서 지원하는 특정기초연구사업으로 수행하였음 (과제번호 : R01-2002-000-00141-0)

2. 시트콤 동영상 분석

시트콤은 장르 자체의 특성상 실내 촬영이 대부분이다. 또한 촬영되는 장소 혹은 배경이 몇 개로 한정되어있다. <그림 1>에서 볼 수 있는 것과 같이 '뉴 논스톱 3'시트콤은 10개의 정형화된 배경이 존재한다. 이러한 배경은 서로 다른 회에서도 거의 비슷하다.



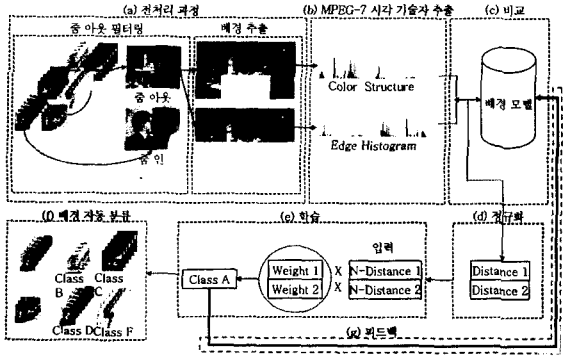
<그림 1> 정형화된 배경

드라마나 영화 같은 동영상은 샷 시퀀스에 있어서 정형적인 특징을 지니고 있지 않다. 그러나 시트콤의 경우 <그림 2>와 같이 줌 아웃 샷과 여러 개의 줌 인 샷이 반복되는 정형성이 존재한다. 이러한 시퀀스가 아닌 샷은 매우 드물기 때문에 무시할 수 있다. 본 논문에서는 시트콤의 이러한 특성들을 이용하여 배경들을 자동으로 분류하게 된다.

<그림 2> '논스톱 3' 동영상에서의 줌 아웃과 줌 인의 반복

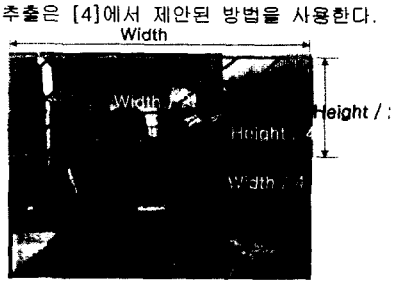
3. LVQ에 기반한 자동 배경 분류 알고리즘

3.1 배경 자동 분류 알고리즘의 전체 구조



<그림 3> 배경 자동 분류 알고리즘

<그림 3>은 본 논문에서 제안한 배경 자동 분류 알고리즘의 전체적인 구조를 나타낸 것이다. 동영상을 샷 단위로 나누는 동영상 인덱싱 과정은 자동 분류 알고리즘 전에 이미 되어 있다고 가정하기 때문에 전처리 과정((a))에서도 제외되어 있다. 먼저, 전처리 과정에서는 샷들을 줌 아웃과 줌 인으로 구별하기 위한 필터링을 실시하여 줌 아웃만을 걸러낸다. 나머지 전처리 과정에서는 걸러진 줌 아웃 샷에서 배경을 추출하는 과정을 거치게 된다. 여기서 배경을 추출하는 이유는 MPEG-7의 시각 기술자를 추출하기 위함이다. 추출 영역은 <그림 4>와 같이 전체 프레임에 대한 것이 중심 부분에 있는 객체들(등장인물들)의 정보를 제외한 배경 정보만을 추출[3]하고 있다. 이러한 배경의 추출은 [4]에서 제안된 방법을 사용한다.

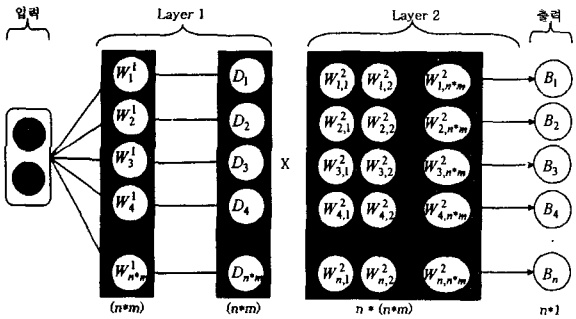


<그림 4> 배경 추출 영역

이렇게 추출된 배경 영역에서 MPEG-7 시각 기술자인 CS(Color Structure)와 EH(Edge Histogram)를 추출((b))한다. 이 두 개의 시각 기술자들은 색과 질감을 대표하며, 같은 배경들끼리는 유사성을 가지며, 각 배경마다 다른 특성을 유지하는 특성을 지니고 있다. 이렇게 추출한 두 가지의 MPEG-7 시각 기술자를 사용하여 각 배경의 모델과 비교((c))하게 된다. 모델과 비교하는 방법은 MPEG-7자체의 비교 방법(Matching Metric)을 사용하게 된다. 이렇게 비교한 결과로서 나오는 입력과 배경 모델의 차이는 이를 정규화하는 과정((d))을 거치게 된다. 정규화를 거쳐 나온 결과를 사용하여 어떤 배경에 속한 것인지를 분류하는 과정을 거치게 된다. 이렇게 분류된 결과를 사용하여 LVQ망은 피드백((g))에 의한 학습((e))을 실시한다. 학습은 올바른 분류를 했을 경우에는, 입력과 비교 대상이 되는 배경 모델을 입력과 가까운 쪽으로 옮기는 과정을 거친다. 이에 반해서 잘못된 분류를 했을 경우에는 입력과 반대되는 방향으로 배경 모델을 움직이게 된다. 배경 자동 분류 알고리즘은 (a)에서 (g)의 과정을 거치게 된다.

3.2 LVQ에 기반한 배경 자동 분류 알고리즘

3.1의 프레임 워크에서 LVQ가 적용되는 사용하는 부분은 Matching, 정규화, 학습, 학습에 의한 자동 분류 그리고 피드백이다. <그림 5>는 본 논문에서 사용하는 알고리즘의 LVQ의 구조이다. 여기서 n은 배경(Background)의 수를 m은 각 배경마다의 클러스터(cluster)의 수를 의미한다. 여기서의 LVQ는 원래의 LVQ와 몇 가지 다른 점을 가지고 있다. 첫 번째로 레이어1의 가중치가 한 종류의 배경을 대표하는 값을 가지는 배경 모델을 의미하게 되는 것이다. 본 논문에서는 이러한 배경을 대표하는 모델이 같은 배경 장면을 갖는 배경들을 하나로 묶는다는 의미로 클러스터로 부르기로 한다. 입력으로 들어온 패턴과 비교하는 방식은 LVQ에서 사용하는 L-1 디스턴스를 계산하는 방식이 아닌, MPEG-7에서 제안한 비교 방식을 사용하게 된다.



<그림 5> 배경자동분류에서 사용되는 LVQ Network

두 번째로 다른 점은 첫 번째 레이어의 클러스터의 변경 방식에 있다. LVQ의 가중치 변경 방법은 들어온 입력이 올바르게 분류되었을 때와 틀리게 분류되었을 때의 두 가지 경우 나눠서 변경되게 된다. 그러나 본 논문에서는 각 클러스터가 하나의 배경 모델을 의미하게 되고, 이러한 배경 모델을 구성하고 있는 CS와 EH가 표현하는 정보의 의미가 다르게 된다. 그러므로, CS와 EH의 변경 방식은 입력된 CS와 가중치를 구성하는 CS 사이의 거리를 구하는 MPEG-7 비교 방법, 입력된 EH와 가중치를 구성하는 EH 사이의 거리를 구하는 MPEG-7 비교 방법을 적용한 새로운 방식을 사용한다. <그림 6>은 위의 내용을 적용한 배경 자동 분류 알고리즘을 의사 코드로 작성한 것이다.

```
Procedure Automatic_Classification_of_Background
Begin
  while( All Clusters Each Backgrounds) do
    Begin
    // n번째 배경의 m번째 클러스터와
    // 입력의 CS와 EH를 계산한다 .
    Dist ( n, m ) =  Dist_of_CS  ( P(CS), W( 1, n, m ) );
                   + Dist_of_EH  ( P(EH), W( 1, n, m ) );
    End
  // Dist 값이 가장 적은 Dist를 찾아낸다
  Winner = Compete( Dist );
  while( All Backgrounds ) do
    Begin
    Find( Winner * W( 2, j ) );
    End
  // 가중치를 update 한다 .
  If Wrong Classification
    UpdateWeight  ( Winner, Wrong ) ;
  Else
    UpdateWeight  ( Winner, Correct );
```

<그림 6> LVQ를 사용한 배경 자동 분류 알고리즘

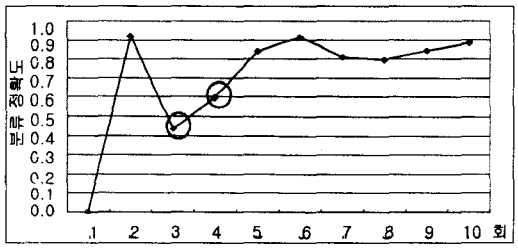
4. 실험 및 결과 분석

4.1 실험 방법

3장에서 제안한 알고리즘을 바탕으로 시트콤 배경의 자동 분류 시스템을 구현하였다. 구현은 MS Windows XP 환경에서 Visual C++6.0으로 하였다. 가장 처음으로, 시트콤 동영상을 샷으로 나누는 방법은 Luminance의 변화를 이용한 방법을 사용하였다[5]. 두 번째로 각 샷을 줌 인과 줌 아웃으로 구분하고, 줌 아웃 샷에서만 키 프레임을 추출하여 배경 정보를 추출한다. 세 번째, 배경으로 추출된 부분에서 MPEG-7 시각 기술자인 CSD와 EHD를 추출한다. 추출은 MPEG-7에서 제안한 방식을 사용하였다. LVQ의 입력은 256개의 bin을 갖는 CSD와 80개의 bin을 갖는 EHD를 사용하였다. 학습시킨 LVQ망을 사용하여, 다른 회의 시트콤을 자동으로 분류한다. 이 결과와 사용자의 피드백을 받아서 다시 LVQ망을 학습시킨다. 이와 같은 학습 과정을 n회까지 실행한다.

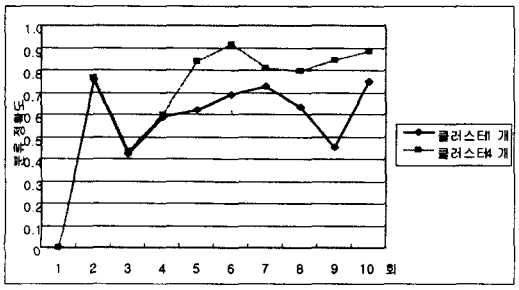
4.2 결과 분석

실험은 서로 다른 두 시트콤 동영상인 '뉴 논스톱 3' 의 10회분과 '똑바로 살아라' 7회분을 사용하였다. 제안하는 알고리즘의 성능은 특정 배경에 관한 샷 중에서 정확하게 분류한 샷의 개수를 해당 배경의 전체 샷 수로 나눈 값을 분류 정확도로 사용하였다. 또한 하나의 시트콤 동영상의 전체 샷 중에서 정확하게 분류한 샷의 개수를 전체 샷 수로 나눈 값을 한 시트콤 동영상에 대한 분류 정확도로 사용하였다.



<그림 7> '논스톱 3' 동영상에서 각 회에 따른 분류 결과

<그림 7은> 각 회에 따른 분류 정확도를 나타낸다. 3회에서는 급격하게 분류 정확도가 떨어지는 모습을 보인다. 이는 아직 학습되지 않은 배경이 입력으로 들어왔기 때문이다. 3회에서 처음으로 입력된 'cafe' 배경(14개의 샷)와 '과방' 배경(16개의 샷) 그리고 '동아리 방' 배경(23개의 샷)이 그 원인이라고 할 수 있다. 그러나 회가 거듭될수록 정확도는 증가하여 0.9로 수렴함을 볼 수 있었다.



<그림 8> 클러스터 수에 따른 배경 분류 결과

<그림 8>은 클러스터 개수를 1개로 한 경우와 4개로 한 경우를 비교한 결과이다. 첫 번째 레이어의 클러스터가 1개일 경우에는 같은 배경에 대해서 다른 입력이 들어오게 되면, 그 배경에 대해서 클러스터가 다시 학습되기 때문에, 그 전에 학습되었던 배경에 대한 클러스터 정보를 잃어버리기 때문에 정확도가 낮다. 그에 반해서 클러스터가 4개일 경우에는 다른 클러스터가 같은 배경의 다른 입력에 대해서 학습되기 때문에, 그 전에 학습되었던 정보를 잃어버리지 않아 점점 분류 성능이 좋아졌다. 그러나 5개 이상일 경우 4개에 비하여 정확도가 그리 증가하지 않았다. 또한, 실험을 통하여 학습 시간이 클러스터의 수에 비례하여 증가하는 것을 알 수 있었다. 따라서 이 실험에서는 적정한 클러스터의 수가 4개임을 알 수 있다.

5. 결론 및 앞으로의 연구 방향

본 논문에서는 동영상 검색에 필요한 부가적인 정보 중에서 시리즈물에 대한 자동 분류 방법을 제안하였으며, 이를 바탕으로 시트콤 동영상 배경에 대한 자동 분류 방법을 구현하였다.. 처음에는 학습되지 않은 새로운 배경이 입력으로 들어옴에 따라서 자동 분류 확률이 떨어지는 모습을 보이지만, 모든 배경이 학습 된 후에는 0.8~0.9의 높은 확률로 수렴되어 가는 모습을 보인다. 이와 같이 본 논문에서는 MPEG-7 시각 기술자들을 사용하여, 시트콤 동영상에서 배경이라는 고급 수준의 정보를 추출하였다.

앞으로의 연구에서는 배경을 자동으로 분류하는 데 있어서 사용된 MPEG-7의 시각 기술자에 대하여, 각각의 배경이 기술자에 있어서 좀 더 명확한 특징을 보이도록 패턴화가 필요하다.

참고 문헌

[1] Martin T. Hagan, Howard B. Demuth and Mark Beale, *Neural Network Design, published by THOMSON LEARNING*, pp.2.2-2.15, 14.2-14.21

[2] 최이정, *시트콤 구조 분석론, 커뮤니케이션 북스*, pp.42, 1999.

[3] N.V.Patel and I.K.Sethi, "Video Shot Detection and Characterization for Video Databases," *Proc. of SPIE Storage and Retrieval for Image and Video Database*, pp.218-225, 1997

[4] Kender, J.R., Boon-Lock Yeo., "Video scene segmentation via continuous video coherence," *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.23-25, 1998.

[5] S. F. Chang, W. Chen and H. J. Meng, "A Fully Automated Content Based Video Search Engine Supporting Spatio-Temporal Queries," *IEEE Transactions on Circuits & Systems for Video Technology*, Vol. 8, pp. 602-615, 1998.