

# 비디오 분석을 위한 자막프레임구간과 자막영역 추출

임 문 철<sup>†</sup> · 김 우 생<sup>††</sup>

## 요 약

최근 하드웨어와 압축기술의 발달로 멀티미디어 데이터에 대한 요구가 급증하고 있다. 비디오 데이터는 비정형으로 되어있고 용량이 커서 내용기반 등 다양한 검색 기법이 요구된다. 비디오에 인위적으로 추가된 자막(Superimposed caption)은 비디오 내용을 분석하는데 중요한 역할을 하며 다양한 검색을 위한 색인 정보로 사용될 수 있다. 본 연구에서는 비디오 프레임 내의 자막영역이 가지고 있는 텍스처 특성을 분석하여 자막영역을 분할하고 프레임들 간에 자막영역의 연속성을 이용하여 정확한 자막프레임구간과 대표자막영역 및 색상을 추출하는 방법을 기술한다.

## Extraction of Superimposed-Caption Frame Scopes and Its Regions for Analyzing Digital Video

Moon-Cheol Lim<sup>†</sup> · Woo-Saeng Kim<sup>††</sup>

## ABSTRACT

Recently, Requirement for video data have been increased rapidly by high progress of both hardware and compression technique. Because digital video data are unformed and mass capacity, it needs various retrieval techniques just as content-based retrieval. Superimposed-caption in a digital video can help us to analyze the video story easier and be used as indexing information for many retrieval techniques. In this research we propose a new method that segments the caption as analyzing texture feature of caption regions in each video frame, and that extracts the accurate scope of superimposed-caption frame and its key regions and color by measuring continuity of caption regions between frames.

### 1. 서 론

하드웨어와 압축기술의 발달로 멀티미디어 데이터를 저장하고 관리하기 위한 요구가 최근 들어 급증하고 있는 추세이다. 가장 대표적인 멀티미디어 데이터인 비디오의 효율적인 검색과 색인을 위해서는 비디오 데이터 내에 포함된 음성, 문자, 영상 등의 정보를 처리할 인식기술, 자연어처리기술, 화상처리 기술들이 필요

하다. 비디오 색인 등 비디오 분석을 위한 한가지 중요한 정보는 비디오 내에 인위적으로 추가된 자막(superimposed caption)이며 비디오의 내용을 이해하는데 중요한 역할을 한다. 자막을 색인정보 등에 이용하기 위해서는 자막 프레임을 감지하고 자막영역을 추출한 후 자막을 인식하는 기술이 필요하다[1].

비디오 데이터를 분석하기 위해 문자를 분할하려는 연구는 기존에도 있었지만 많은 제약조건을 가지고 있었다. R. Lienhart의 [2, 3]에서는 비디오 프레임에 출현하는 문자 중에서 인위적 문자(artificial text)에 관심을 두고 우선 분리과 합병(Split-and-Merge) 알고리

※ 본 논문은 2000년도 교내학술 연구비 지원에 의한 결과임.  
† 준 최 원 · 광운대학교 대학원 컴퓨터공학부  
†† 종신희원 · 광운대학교 컴퓨터공학부 교수  
논문접수 : 2000년 9월 6일, 심사완료 : 2000년 11월 13일

즘으로 영역들을 분할한 다음 배경영역을 제거하기 위해 문자의 최대와 최소크기 값을 사용하였다. 이들은 동일한 색상영역을 하나의 객체로 분할하는데 잡음이 많은 비디오 프레임에서 무수히 많은 객체들이 분할되므로 문자객체와 배경객체의 구분이 어려웠다. 따라서, 분리와 합병 알고리즘 대신 영역성장(Region Growing) 알고리즘과 합병기(Merger)를 사용하여 이러한 문제를 해결하려고 하였으며 각 객체를 프레임마다 동적으로 추적하여 생성되거나 사라진 객체를 연결 또는 제거함으로써 각 문자객체를 분할하였다[4]. 그러나, 이러한 방법도 자막문자의 크기에 민감하고 많은 객체들을 추적하여 분할하는 알고리즘의 비용이 많이 소요된다. 또한, 자막문자의 수직 에지 특징(Vertical edge features)을 기반으로 하여 이러한 에지를 결합하고 결합된 에지 영역의 크기와 밀집도 및 수평-수직 크기 비율에 임계값을 적용하여 자막영역을 분할하는 연구도 있었다[5, 6]. 이는 복잡한 배경의 비디오에서 자막을 구성하는 에지 영역과 배경 에지 영역의 구분이 어렵고 자막색상의 제한과 각 임계값의 크기에 매우 민감하며 뉴스나 특정한 비디오 데이터를 실험대상으로 하였다. A. K. Jain의 [7]에서는 하나의 입력 영상으로부터 컬러 레벨에 따라 여러 개의 전경영상을 만들고 객체의 연결 구성요소 분석(connected component analysis)을 통하여 객체들을 연결함으로써 문자영역을 추출하는데 복잡한 영상에서 문자와 배경이 비슷한 컬러 레벨로 근접되어 있을 때 정확한 문자영역을 추출하지 못하는 문제점이 존재한다. 일반적으로 비디오 데이터는 배경이 매우 복잡하고 자막문자의 색상과 크기가 다양하기 때문에 자막색상과 자막문자의 크기에 민감하지 않고 복잡한 배경에서도 자막영역을 정확히 분할할 수 있는 방법이 요구된다.

본 연구에서는 비디오 프레임에 인위적으로 첨부된 자막의 색상 단일성과 텍스트 특성을 분석하여 자막색상과 크기에 상관없이 복잡한 배경에서도 자막영역을 정확하게 분할하는 방법을 기술하고 동일한 자막들이 일련의 프레임 집합들에 연속적으로 포함되어 있으며 그 위치와 크기 및 색상이 거의 변하지 않는다는 특성을 이용하여 동일자막 프레임 집합인 지딕프레임구간과 대표자막영역 및 색상을 추출하는 방법을 제안한다. 본 논문은 각 프레임의 자막영역을 분할한 후 이들에 대한 위치와 크기 및 색상의 유사성을 평가하고 연속성을 측정하여 비교적 정확한 자막프레임구간과

자막영역을 추출할 수 있음을 실험을 통하여 보였다.

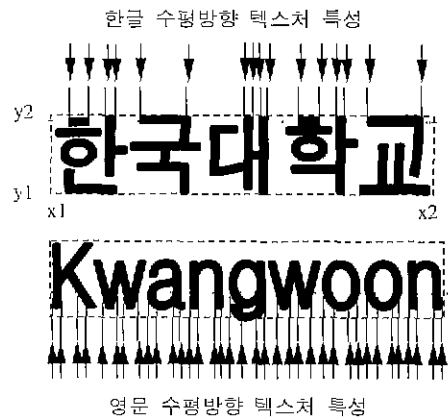
본 논문은 2장에서 자막영역에서의 기울기 특징 및 색상의 유사성을 이용하여 단일 프레임에서 자막영역을 분할하는 방법을 기술한다. 3장에서 동일자막이 연속된 프레임에 존재하는 특징을 이용하여 자막프레임 구간을 결정한 후 자막영역을 추출하는 방법을 기술한다. 4장에서 실험 및 결과분석을 5장에서 결론 및 향후과제에 대해 기술한다.

## 2. 자막영역 분할

비디오의 복잡한 배경으로부터 자막영역을 분할하기 위해 우선 일관 문자열의 특성을 살펴본 후 자막영역을 배경과 현격하게 구별할 수 있는 특징을 추출하여 각 프레임의 자막영역을 배경으로부터 분할하는 과정을 기술한다.

### 2.1 특징 추출

자막영역은 대부분 하나 이상의 문자열로 존재하며 문자열의 수평 방향으로 단일색상이 주기적(periodic)으로 존재하여 배경과 뚜렷하게 구분되는 특성이 있다. 즉, 프레임의 가 수평라인에 대해서 X축 방향인 왼쪽에서 오른쪽으로 자막과 배경 색상의 큰 차이가 주기적으로 반복되는 맥스처 특성이 존재한다. (그림 1)은 임의의 한글과 영문 문자열에 나타나는 수평방향의 텍스처 특성을 보여주고 있다.



(그림 1) 수평방향의 자막영역 특성

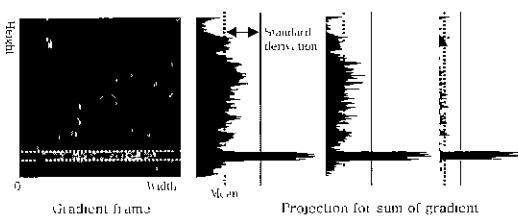
위와 같은 맥스처 특성을 기반으로 [8]에서는 각 문자의 색상과 배경의 색상 차이가 크게 나타나는 부분

이 일정한 거리 내에서 주기적으로 반복되는 개수인 색상변화 빈도수를 이용하여 각 프레임에서 자막영역을 분할하였다. 그러나, 일정한 거리와 색상변화 빈도수에 대한 임계값 설정 문제와 문자열이 RGB 각 색상만으로는 배경과 구분되지 않을 수 있는 문제를 보완하기 위해 본 연구에서는 색상변화 빈도수 대신에 문자열 부분의 기울기 특징을 이용하여 배경으로부터 문자열을 분할하는 방법을 사용한다.

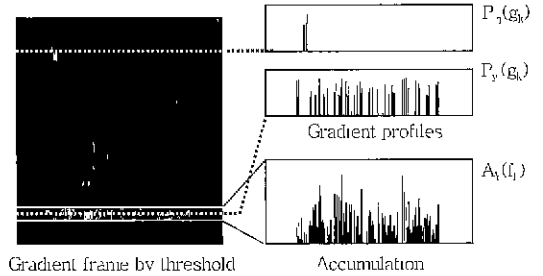
비디오의 각 프레임은 매우 복잡한 모양과 색상으로 구성된 배경을 가지고 있으며 심지어 동일한 색상으로 구성되어 있어야 할 자막의 색상도 24bit RGB 색상에서는 다양한 색심 값으로 구성되어 있다 따라서, 기울기 특징을 추출하기 전에 자막의 형태가 변경되지 않는 범위 내에서 복잡한 배경과 자막색상을 단순화하여 자막영역 추출 오류를 최소화하기 위한 방법으로 우선 각 프레임의 RGB 색상에 식 (1)을 적용한다[9].  $D(x, y)$ 는 각 픽셀의 색상축소 결과,  $O(x, y)$ 는 각 픽셀의 원래색상,  $K$ 는 원래 색상의 수,  $G$ 는 축소하고자 하는 색상의 수를 의미한다.

$$D(x, y) = \lfloor \frac{G \times O(x, y)}{K} \rfloor \times \frac{K}{G} \quad (1)$$

자막 특징은 색상이 단순화 된 RGB 각 프레임에 대한 각 수평 라인 화소들의 기울기 측면도(profiles)와 각 라인마다 기울기 크기의 합으로 표현할 수 있다 기울기 크기는  $x$  방향 소벨 마스크  $G_x$ 에 의해 구성되며 기울기 크기의 합은 수평방향에 투영(projection)하여 자막 영역과 비 자막영역의 특징을 구분할 수 있다. (그림 2)는 기울기 프레임에서 기울기 크기의 합에 대한 투영도를 보여주고 있으며 임계값을 적용한 세계의 기울기 투영도 모두 자막영역이 비 자막영역에 비해 현저히 높은 기울기 크기의 합을 가지고 있다는 점을 알 수 있다 또한, (그림 3)은 임의의 임계값을 적용한 기울기 프레임에서 수평 라인 각 화소에 대한



(그림 2) 기울기 투영도



(그림 3) 기울기 측면도

기울기 측면도를 보여주고 있으며 자막영역의 기울기 값 빈도가 비 자막영역에 비해 높고 축적된 값에서는 현격한 차이가 존재한다는 점을 알 수 있다

### 2.2 자막영역 분할

기울기 투영도와 기울기 측면도를 분석하여 직사각형의 자막영역들을 분할한다. 기울기 투영도는 임계값 이상의 기울기 합이 자막문자열 구간에서 수직방향으로 연속성을 가지기 때문에 자막의 세로인  $Y$ 축 영역 구간을 결정하는데 사용되며 결정된 자막영역의 세로 범위 내에 존재하는 화소에 대한 기울기 측면도는 자막문자열 구간에서 빈도수가 높기 때문에 자막의 가로인  $X$ 축 영역구간을 결정하는데 사용된다.

자막영역은 각 프레임을 2차원  $(x, y)$  좌표로 보았을 때 직사각형의 왼쪽 하단  $(x_0, y_0)$ 과 오른쪽 상단  $(x_1, y_1)$ 로 표현된다. 세로영역  $y_0$ 와  $y_1$ 를 결정하기 전에 우선 (그림 2)의 기울기 투영도에 나타낸 평균과 표준편차를 다음 식들에 의해 계산한다.  $width$ 의  $height$ 는 프레임의 가로와 세로 크기,  $G_x$ 는 소벨 수직 방향 기울기 마스크,  $G_y$ 는 프레임 세로 각 라인의 기울기 크기의 합,  $G_m$ 은  $G_y$ 에 대한 평균,  $G_\sigma$ 는 표준편차이다.

$$G_y = \sum_{x=1}^{width-1} \sqrt{G_x^2} \quad (2)$$

$$G_m = \frac{\sum_{y=1}^{height-1} \sum_{x=1}^{width-1} (\sqrt{G_x^2})}{height-2} \quad (3)$$

$$G_\sigma = \sqrt{\frac{\sum_{y=1}^{height-1} G_y^2 - G_m^2}{height-2}} \quad (4)$$

다음으로  $[G_m + G_\sigma < G_t]$ 를 만족하는  $G_y$ 의 값을 탐색한 후 연속적인 구간을 찾아내어  $(y_0, y_1)$ 의

쌍들을 자막영역의 세로인 Y축 영역구간으로 결정한다. 또한, 결정된  $y_1$ 와  $y_2$  범위 내에서 누적한 기울기 측면도  $A_Y$ 를 식 (5)와 같이 정의하고 가로영역을 결정하는데 사용한다.  $A_Y(f_k)$ 는 누적한 기울기 측면도,  $P_Y(g_k)$ 는 각 수평라인의 기울기 측면도,  $f_k$ 와  $g_k$ 는 각각 누적한 기울기 측면도와 기울기 측면도의 각 기울기 크기를 의미한다.

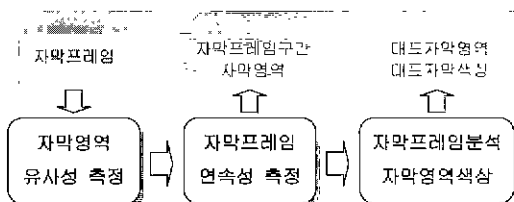
$$A_Y(f_k) = \sum_{y=y_1}^{y_2} P_Y(g_k) \quad (k=0,1,\dots,width) \quad (5)$$

가로영역  $x_1$ 와  $x_2$ 는 누적한 기울기 측면도에서  $f_k$  값들 사이의 거리가 임계값 이하이고 연속된 구간을 자막영역의 가로인 X축 영역구간으로 결정한다.

위와 같이 선택된 세로영역과 가로영역이 매정일 가능성을 배제할 수 없다. 선택된 자막 후보 영역내의 색상이 다양하게 골고루 분포되어 있다면 매정일 가능성이 크며 단일색상으로 분포되어 있다면 자막영역일 가능성이 크다. 따라서,  $(y_2-y_1)$ 와  $(x_2-x_1)$ 로 선택된 각 영역에 대하여 색상별 총 화소 수를 계산하여 최대 값을 가진 색상과 색상별 기울기 크기의 수를 계산하여 최대 값을 가진 색상이 같으면 자막영역, 아니면 배경으로 판단한다.

### 3. 자막프레임구간 및 대표자막영역 추출

비디오 데이터에서 동일한 자막은 임의의 자막프레임구간에서 위치와 크기 및 색상이 일정하며 연속적으로 존재하는 특징이 있다. 본 정에서는 각 프레임에서 미리 분할된 자막영역의 위치와 크기 및 자막색상의 유사성을 비교하여 동일한 자막영역인지를 판단하고 이러한 영역들을 포함하고 있는 자막프레임의 연속성을 측정하여 자막프레임구간 및 대표자막영역을 추출하는 방법에 관하여 기술한다. (그림 4)는 이러한 과정을 도식화하여 보여주고 있다.



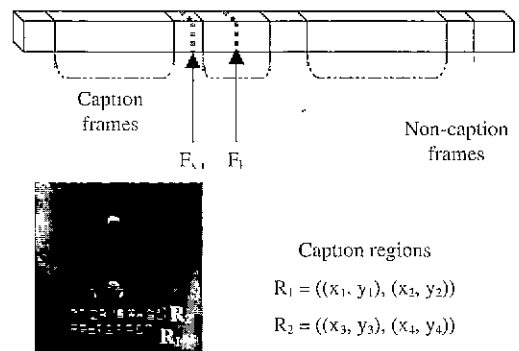
(그림 4) 자막프레임구간 및 대표자막영역 추출 과정

### 3.1 분할된 자막영역의 유사성 측정

분할된 자막프레임들 간의 자막영역들이 동일한 자막영역인지 판단하기 위해 모든 영역들에 대해 위치와 크기 및 색상을 비교하여 유사성을 측정한다. 각 비디오 데이터에서 임의의 분할된 자막영역  $R_n$ 은 직사각형 좌표로 표현되고 임의의 분할된 자막프레임  $F_k$ 는 N개의 분할된 자막영역들을 포함하고 있으며 식 (6)과 같이 표현할 수 있다. (그림 5)는 일련의 연속적인 프레임에 존재하는 자막프레임들과 비 자막프레임들 및 분할된 자막영역들을 보여주고 있다.

$$F_k = \{ R_1, R_2, \dots, R_n, \dots, R_N \} \quad (6)$$

$$R_n = ((x_{2n-1}, y_{2n-1}), (x_{2n}, y_{2n}))$$



(그림 5) 자막프레임과 자막영역

분할된 두 개의 자막프레임들 간에 존재하는 자막이 동일한 자막이라 하더라도 측정된 위치와 크기 및 색상에서 약간의 차이가 있을 수 있으므로 이러한 차이에 영향을 받지 않고 유사성을 판단할 수 있는 방법을 이용한다. 따라서, 이전프레임  $F_{k-1}$ 의 분할된 자막영역을  $R_n = ((x_{2n-1}, y_{2n-1}), (x_{2n}, y_{2n}))$ 이라 하고 현재프레임  $F_k$ 의 분할된 자막영역을  $R_n' = ((x_{2n-1}', y_{2n-1}'), (x_{2n}', y_{2n}'))$ 이라 할 때 두 영역  $R_n$ 과  $R_n'$ 에 대해 다음의 세 가지 조건을 만족하면 두 영역은 유사하며 동일한 자막영역으로 판단한다

- (1) 위치유사성 판단조건  $x_{2n-1} < x_{2n-1}'$  and  $x_{2n} > x_{2n-1}', y_{2n-1} < y_{2n-1}'$  and  $y_{2n} > y_{2n-1}'$
- (2) 크기유사성 판단조건 .

$$\frac{S(D_{R_n, R_n'})}{S(R_n)} > T \text{ and } \frac{S(D_{R_n, R_n'})}{S(R_n')} > T$$

(단,  $D_{R_n, R_n'} = ((\max(x_{2n-1}, x_{2n-1}'), \max(y_{2n-1}, y_{2n-1}'))$

$$(\min(x_{2n}, x_{2n}'), \min(y_{2n}, y_{2n}'))$$

(3) 색상유사성 판단조건  $C_{r,g,b}(R_n') = C_{r,g,b}(R_n)$

$S(R_n)$ 와  $S(R_n')$ 은 각각  $R_n$ 과  $R_n'$ 의 면적이고  $SD_{min}(R_n')$ 은  $R_n$ 과  $R_n'$ 가 겹치는 부분의 면적으로 모두 사각형 내의 총 픽셀 수를 나타내며  $C_{r,g,b}(R_n)$ 와  $C_{r,g,b}(R_n')$ 은 각각 RGB 프레임의  $R_n$ 과  $R_n'$ 에서 최대 화소 수를 가진 색상이다. (1)은 두 영역의 세로와 가로 위치가 서로 겹쳐지는지, (2)는 두 영역의 겹치는 비율이 임계값 T 이상인지, (3)은 두 영역에서 최대 화소 수를 가진 색상이 동일함을 판단한다.

### 3.2 자막프레임구간과 대표자막영역 추출

2장에서 분할된 각 자막프레임은 하나 이상의 자막영역을 포함하게 되는데 첫째, 프레임들 사이에 위치와 크기 및 색상이 유사한 자막영역이 최소한 하나라도 존재하고 서로 대응하는 자막영역의 수가 같은지 둘째, 이러한 자막프레임이 연속적으로 N 개 이상 존재하는지를 측정하여 두 가지 조건을 만족하면 그 구간의 모든 프레임들이 동일한 자막영역을 포함하고 있는 것으로 간주되고 하나의 자막프레임 구간으로 결정한다 만일, 연속적인 자막프레임의 수가 N 개 미만이라면 그 프레임들은 오류 자막프레임으로 간주되며 프레임 내의 자막영역들은 자막영역에서 제외된다. 자막프레임구간을 결정하는 과정은 다음과 같다

- (1)  $F_k$ 가  $F_{k+1}$ 과 연속적인 프레임인지 판단한다.  
연속적이지 않으면, (4)를 수행한다.
- (2)  $F_k$ 과  $F_{k+1}$ 의 자막영역이 동일인지 판단한다 (3.1절의 유사성 조건을 만족하는지 판단)  
동일하지 않으면, (4)를 수행한다.
- (3)  $F_k$ 를 현재 자막프레임구간에 할당하고  $k = k+1$ 한 후 (1)을 수행한다
- (4) 현재 자막프레임구간의 프레임수가 N 이상이면, 자막프레임구간으로 할당한다.  
현재 자막프레임구간의 프레임수가 N 미만이면, 현재 자막프레임구간을 무시한다.
- (5) 다음 자막프레임이 존재하면 (1)을 수행하고, 존재하지 않으면 종료한다.

자막프레임구간을 결정할 때 임계값 N의 크기는 비디오 데이터의 초당 프레임 비율에 따라 달라질 수 있으며 본 연구에서는 4장에서 기술한 바와 같이 일반적

인 비디오 데이터의 분석결과에 따라 결정하였다. 위의 과정을 수행하여 결정된 각 자막프레임구간은 비디오 데이터에서 임의의 연속된 프레임 구간을 차지하며 그 구간 내의 각 프레임들은 동일한 수의 유사한 자막영역들을 포함하게 된다.

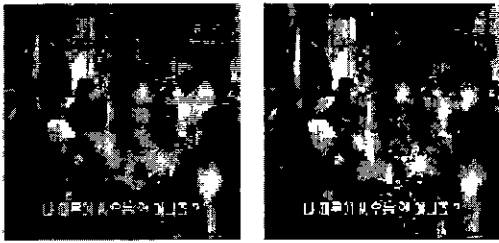
하나의 결정된 자막프레임구간에 속하는 모든 프레임에는 동일한 수의 유사한 자막영역이 존재하므로 자막영역을 분석하여 응용할 때 그 구간을 대표하는 하나의 대표자막영역 만을 사용하면 된다. 또한, 대표자막색상은 자막영역 내에서 배경색상을 배제한 자막색상 만의 분석이 가능하게 되어 자막영역의 각 문자를 정확하게 분리하는데 중요한 정보로 이용될 수 있다 따라서, 본 연구에서는 자막프레임구간의 대표자막영역과 대표자막색상을 결정한다. 추출된 자막프레임구간 내의 자막영역들은 모두 유사하므로 그 구간에서 중간프레임의 자막영역들을 대표자막영역으로 선택하는데 각 영역은  $((x_n, y_n), (x_{n+1}, y_{n+1}))$ 로 표현될 수 있다. 또한, 대표자막영역에서 최대 화소 수를 가지는 색상을 대표자막색상으로 결정한다.

## 4. 실험 및 결과 분석

본 연구에서는 초당 프레임수가 10(10 frames/sec)인 약 3분 길이의 비디오 데이터 1편과 약 5분 길이의 비디오 데이터 5편을 실험에 사용하였는데 이들은 한글자막 비디오 데이터 3편, 영문자막 비디오 데이터 3편이다. 한글자막 3편과 영문자막 1편은 일반 영화 비디오 데이터이며, 영문자막 2편은 자막색상이 노란색인 만화 비디오 데이터이다 또한, 본 논문에서의 주요 처리과정인 자막영역 분할과 자막프레임구간 추출은 초당 8 프레임 속도의 실시간으로 처리되었다.

### 4.1 실험

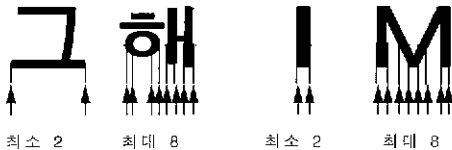
본 연구에서는 2.1절에서 설명한 식 (1)의 G 값을 4로 하여 RGB 각기를 4 계조수로 축소함으로써 결과적으로 각 프레임의 전체 색상을 64(4\*4\*4) 색상으로 축소하였다. (그림 6)의 (a)는 24bit RGB 색상의 원래 프레임이고 (b)는 색상을 축소한 프레임을 보여주고 있는데 배경은 색상이 현저히 축소되었고 자막의 형태는 변하지 않되 더욱 단순화된 색상으로 존재한다는 것을 알 수 있다.



(a) 원래 프레임 (b) 색상축소 프레임  
(그림 6) 원래 프레임과 색상축소 프레임

자막이 배경과는 달리 대부분 전경(foreground)이고 단일색상이라는 특징 때문에 각 프레임의 색상 수를 전체적으로 축소시키 복잡한 배경과 자막색상은 단순히 회색이지만 자막의 형태는 거의 변경되지 않았다.

또한, 자막의 특징추출 시 수평방향 기울기 크기의 합은 자막문자의 개수에 의존하는데 한글 1문자의 경우 최소 2개('그')에서 최대 8개('해')이며, 영문 1문자의 경우 최소 2개('I')에서 최대 8개('M')이며 대부분 4개이다. (그림 7)은 한글과 영문에서 1문자에 대하여 기울기 크기가 증가할 수 있는 부분의 예를 보여주고 있다.



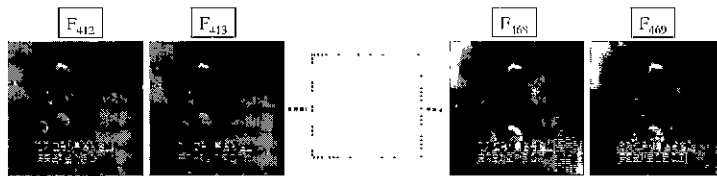
(그림 7) 한글 및 영문에서 기울기가 큰 부분

(그림 8)은 임의의 고정된 자막프레임구간과 그 구간을 대표하여 추출된 대표자막영역 및 대표자막색상을 보여주고 있다. 상단은 임의의 자막프레임구간  $F_{412} \sim F_{439}$  에 속하는 프레임들로 각각 두 개의 자막영역들을 포함하고 있으며 하단의  $F_{441}$ 은 이 구간의 중간프레임으로 대표자막영역의 위치를 시각적으로 보여주고 있다.

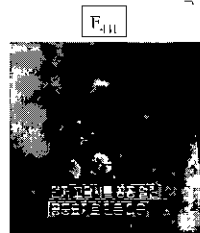
4.2 결과 분석

<표 1>은 비디오 데이터의 자막을 언어별, 색상별로 구분하여 단일 프레임에서 색상변화 빈도수를 이용한 자막영역 예측 실험결과와 자막의 프레임간 연속성을 적용한 자막프레임구간 추출 실험결과를 보여주고 있다.

<표 1>에서  $N_{cp}$ 의 각 측정치들은 총 자막프레임 중에서 자막영역 분할 프레임 수와 백분율, 총 프레임 중에서 배경인데 자막으로 잘못 추출된 배경오류 프레임 수와 백분율이다.  $N_{cp}$ 에서 분할 프레임 수는 자막영역을 포함하고 있는 프레임들 중에서 3 문자 이상의 자막영역을 하나 이상 포함하고 있는 것으로 판단하여 분할된 프레임이며, 자막영역이 존재하는데 분할되지 않은 프레임들은 대부분 3~5 문자로 구성된 자막들을 포함하고 있으며 비디오 데이터의 화질로 인해 문자가 정확하게 보이지 않는 경우가 많았다. 또한, 오류 프레임 수는 자막이 존재하지 않지만 배경이 자막영역으로 잘못 판단된 경우인데 이들은 자막과 비슷한 텍스트



자막프레임구간 번호 : 3 (F412 ~ F469)  
자막프레임구간의 프레임 수 : 58  
각 프레임의 자막영역 : 2



대표자막 영역  
영역  $r_1$  (43, 20), (139, 31)  
영역  $r_2$  (44, 38), (151, 49)  
대표자막 색상  
R (192)  
G (192)  
B (192)

자막프레임구간의 중간프레임

(그림 8) 자막프레임구간과 대표자막영역 및 대표자막색상 추출

<표 1> 자막영역 예측과 자막프레임구간 추출 실험결과

비디오 데이터	자막 언어	자막 색상	총 프레임 수	총 자막 프레임 수	총 자막 구간 수	단일프레임 기반 자막영역 분할(Ncp)		자막의 연속성에 의한 자막프레임구간 추출(Nfs)		
						분할 프레임 수 (%)	배경 오류 수 (%)	추출자막 구간 수 (정확)	추출자막 구간 수 (누적)	배경 오류 수 (%)
video1	한글	흰색	2819	785	21	770 (98.09)	20 (0.70)	20	1	0 (0)
video2	한글	흰색	2882	930	27	913 (98.17)	25 (0.85)	26	1	0 (0)
video3	한글	흰색	1817	772	19	761 (98.57)	18 (0.99)	18	1	0 (0)
video4	영문	흰색	2710	1987	28	1933 (97.28)	33 (1.21)	26	2	0 (0)
video5	영문	노랑	2792	1205	30	1155 (95.85)	42 (1.50)	28	2	0 (0)
video6	인문	노랑	2745	1324	30	1270 (95.92)	41 (1.60)	28	2	0 (0)

특성을 가지며 대부분 프레임들에 연속적으로 나타나지는 않았다

Nfs의 각 측정치들은 모든 자막영역이 정확하게 추출된 자막구간 수(정확), 몇 개의 자막영역이 누락된 자막구간 수(누락), 총 프레임 중에서 배경인데 자막으로 잘못 추출된 배경오류 프레임 수와 백분율이다. Nfs에서 자막구간은 모두(정확+누락) 추출되었는데 이는 단일 프레임에서 분할된 자막영역들이 동일 자막구간의 각 프레임에서 유사한 크기로 정확하게 분할되었고 실제의 자막영역들이 분할된 자막영역들에 모두 포함되어 있었기 때문이다 또한, 배경 오류 수는 모두 0인데 이는 자막영역 분할에서 발생했던 배경 오류가 일련의 프레임에서 동일하지 않고 연속적으로 존재하지 않기 때문에 자막의 동일성과 연속성을 이용함으로써 모두 제거되었다는 점을 알 수 있다.

Nfs는 분할된 자막영역의 유사성과 연속성을 이용하여 추출되었는데 유사성 판단을 위해 위치와 크기 및 색상의 유사성 판단조건을 사용하였으며, 세 가지 중에서 크기 유사성 판단 시에는 임계값이 필요한데 자막영역의 면적이 90% 이상 겹치면 동일한 영역으로 판단하였다. 또한, 연속성을 이용한 자막프레임구간 추출을 위해 일반적인 비디오 데이터에서 동일한 자막이 10 개의 프레임에서 연속적으로 존재하면 동일 자막프레임구간으로 결정하였는데 이는 초당 10 프레임(10 frames/sec)인 일반 비디오 데이터로부터 동일자막이 연속적으로 존재하는 프레임의 개수를 측정할 결과로부터 최소값을 선택한 것이며 초당 프레임의 개수가 더 많은 비디오 데이터에서는 10 보다 더 큰 측정치를 얻을 수 있을 것이다.

5. 결 론

본 연구에서는 단일 프레임에서 자막의 텍스트 특성

인 기울기 투영도와 측면도를 이용하여 자막영역들을 분할하였으며 분할된 자막영역들의 위치와 크기 및 색상의 유사성을 평가하고 프레임들 사이에서 이러한 영역들의 연속성을 측정하여 실제 동일한 자막이 존재하는 자막프레임구간과 비 자막영역을 제거하고 자막의 문자분리 및 인식에 필요한 대표자막영역 및 대표자막색상을 추출하는 방법에 관하여 기술하였다. 일련의 프레임 구간에서 자막의 동일성과 연속성을 이용함으로써 단일 프레임에서 자막과 배경을 구분하기 어려워 발생한 배경오류 자막을 제거하여 자막영역 추출의 신뢰성을 향상시킬 수 있고 비디오 전체 프레임의 구조 분석에 필요한 자막프레임구간을 비교적 정확하게 추출할 수 있음을 실험결과로 보여주었다. 본 연구는 가능한 정확하게 자막프레임구간을 추출하고 그 구간에 대한 대표자막영역 및 대표자막색상을 추출하는데 초점을 두었기 때문에 추후 자막을 인식하고 인식한 내용과 자막프레임구간 정보를 이용하여 인덱싱하는 방법에 대한 연구가 요구된다.

참 고 문 헌

[1] H D. Wactlar, T Kanade, M. A Smith, and S M Stevens, "Intelligent access to digital video : The Informedia ptoject," *IEEE Computer*, Vol 29, No.5, pp.46~52, May, 1996.

[2] R Lienhart and F. Stuber, "Automatic Text Recognition in Digital Videos," *In Image and Video Processing IV 1996, Proc SPIE 2666-20*, January 1996

[3] R Lienhart, "Automatic Text Recognition for Video Indexing," *In Proceedings of the ACM Multimedia 96, (Boston, MA, USA, November 11-18, 1996)*, S

11-20, November 1996.

[4] R. Lienhart, "Automatic Text Segmentation and Text Recognition for Video Indexing," *ACM/Springer Systems Magazine*, September 1998.

[5] M. A. Smith and T. Kanade. "Video Skimming and Characterization through The Combination of Image and Language Understanding Technique." In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 775~781, 1997

[6] Toshio Sato, Takeo Kanade, Ellen k Hughes, Michael A. Smith, "Video OCR for Digital News Archives." *IEEE International Workshop on Content-Based Access of Image and Video Database (CAIVD '98)*, 1998.

[7] A. K. Jain and B. Yu, "Automatic text location in images and video frames." *Technical Report MSUCPS : TR97-33. Dept of Computer Science, Michigan State University*, 1997.

[8] 임문철, 김우생. "비디오에서 색상변화 빈도수를 이용한 자막영역 추출기법". *한국 멀티 미디어 학회 '99 춘계 학술발표 논문집*, 제2권 1호, pp 121-126, 1999

[9] Earl Jost, "Pattern recognition and image analysis," Prentice Hall Inc., pp 273, 1996



### 임 문 철

e-mail mclim@cs.kwangwoon.ac.kr  
 1994년 순천대학교 전자계산학과 졸업(학사)  
 1996년 광운대학교 대학원 전자계산학과(이학석사)  
 1996년~현재 광운대학교 컴퓨터 공학부 박사과정

관심분야 : 영상/미디어 처리, 패턴인식, 컴퓨터비전



### 김 우 생

e-mail woosaeng@cs.kwangwoon.ac.kr  
 1982년 서울대학교 수료  
 1985년 University of Texas at Austin 졸업(학사)  
 1987년 University of Minnesota (이학석사)

1987년~1988년 현대전지, Zeus Computer 과장  
 1991년 University of Minnesota (이학박사)  
 1992년~1996년 광운대학교 전자계산학과 조교수  
 1996년~현재 광운대학교 컴퓨터공학부 부교수  
 관심분야 : 멀티미디어, 영상/미디어 처리 및 패턴인식