

# 일반화된 문자 및 비디오 자막 영역 추출 방법

## (A Generalized Method for Extracting Characters and Video Captions)

전 병 태 <sup>†</sup> 배 영 래 <sup>\*\*</sup> 김 태 윤 <sup>\*\*\*</sup>

(Byung Tae Chun) (Younglae Bae) (Tai-Yun Kim)

**요약** 기존의 문자 영역 추출 방법은 전체 영상에 대하여 컬러 축소(color reduction), 영역 분할 및 합병(region split and merge), 질감 분석(texture analysis) 등과 같은 방법을 이용하여 문자 영역을 추출했다. 이 방법들은 많은 휴우리스틱(heuristic) 변수와 추출하고자 하는 문자의 사전 지식에 의해 임계치 값을 설정함으로써 알고리즘을 일반화하기 어렵다는 문제점이 있다.

본 논문에서는 문자의 지형학적 특징점 추출 방법과 점-선-면 확장법을 이용하여 문자 영역을 추출함으로써 기존 문자 영역 추출의 문제점인 휴우리스틱 변수의 사용을 최소화하고 임계치 값을 일반화함으로써 일반화된 문자 영역 추출 방법을 제안 하고자 한다. 실험결과 일반화된 변수와 임계값을 사용함으로써 문자의 사전 지식 없이도 문자 영역을 추출함을 볼 수 있었다. 비디오 영상의 경우 후보 영역 추출율 100%, 검증을 통한 자막 영역 추출율은 98% 이상임을 볼 수 있었다.

**Abstract** Conventional character extraction methods extract character regions using methods such as color reduction, region split and merge and texture analysis from the whole image. Because these methods use many heuristic variables and thresholding values derived from a priori knowledge, it is difficult to generalize them algorithmically.

In this paper, we propose a method that can extract character regions using a topographical feature extraction method and a point-line-region extension method. The proposed method can also solve the problems of conventional methods by reducing heuristic variables and generalizing thresholding values.

We see that character regions can be extracted by generalized variables and thresholding values without using a priori knowledge of character region. Experimental results show that the candidate region extraction rate is 100%, and the character region extraction rate is over 98%.

### 1. 서론

최근 컴퓨터 하드웨어의 급속한 발전으로 처리 속도 향상과 메모리의 대용량화가 가능하게 됨으로서 과거 컴퓨터 환경의 제한으로 개발되지 못했던 많은 기술들의 개발이 가능하게 되었다. 컴퓨터 환경의 향상과 멀티미디어 기술의 급속한 발전으로 대용량 멀티미디어 저

장/서버 기술이 발전하게 되었으며 그 기술 중 하나가 영상/동영상 데이터 베이스 기술이다.

기존의 영상/동영상 데이터 베이스 시스템은 단순한 키워드에 의하여 검색함으로써 다양한 사용자의 검색 요구에 부응할 수 없는 단점이 있었다. 이러한 문제점을 해결하기 위하여 최근에 내용기반 동영상 검색 시스템 개발에 많은 관심을 갖게 되었다. 내용 기반 동영상 검색 시스템은 추출된 영상의 특징 정보(컬러, 모양, 자막, 텍스트, 건본 영상, 등)를 메타 데이터화하여 데이터 베이스에 저장하고, 이 특징에 대하여 질의함으로써 질의 특징과 가장 유사한 영상을 추출하여 주는 시스템이다 [1][2][3]. 이들 특징 정보 중 영상의 의미를 가장 함축 적으로 표현해놓은 의미 정보는 비디오에 존재하는 자막 정보일 것이다. TV 프로그램에서 뉴스, 여행안내,

<sup>†</sup> 정 회 원 : 한국전자통신연구원 컴·소·연 영상처리연구부 연구원  
chunbt@etri.re.kr

<sup>\*\*</sup> 비 회 원 : 한국전자통신연구원 컴·소·연 영상처리연구부 연구원  
yijb@etri.re.kr

<sup>\*\*\*</sup> 종신회원 : 고려대학교 컴퓨터학과 교수  
tykim@netlab.korea.ac.kr

논문접수 : 1999년 9월 14일

심사완료 : 2000년 2월 16일

취미, 교육 훈련에 관한 문자 정보는 대표적인 영상의 의미 정보가 될 수 있다. 따라서 이러한 문자 정보를 자동적으로 추출하고 인식하여 의미 화할 수 있는 연구가 필요하다.

문자 영역 추출에 관한 기존의 연구[4][5][6]는 공장 자동화, 사무 자동화, 산업 자동화에서 필요로 하는 문자 추출 및 인식에 관한 연구가 주로 이루어 졌다. 이들 시스템은 어느 정도 좋은 연구 결과가 이루어져 여러 분야에서 일부 사용되고 있다. 그러나 이 연구들은 문서 영상과 같은 이진 영상을 처리하기 위한 방법으로 명암도(gray value)나 컬러 영상과 같은 영상을 처리하기에는 부적합하다고 볼 수 있다.

기존의 문자 영역 추출 방법은 자막을 텍스처로 간주하고 가보 필터링[7]과 같은 텍스처적 분석 방법[8]을 이용하거나 공간적 변화 분석 방법[9]을 이용하여 자막 영역을 추출하였다. 위와 같은 텍스처 분석 방법은 문자의 크기 및 서체에 민감하고 문자 영역을 정확히 추출할 수 없을 뿐 아니라 처리 시간이 많이 소요된다는 단점이 있다.

전체 영상에 대하여 영역 분할 및 합병(region split and merge) 방법[10]을 이용하여 문자 영역을 추출하거나, 프레임간의 차를 이용한 문자 영역 추출 방법[11]들은 문자에 대한 사전 지식이 필요하며 전체 영상을 처리함으로써 처리 시간이 많이 소요된다는 단점이 있다.

컬러의 축소(color reduction) 방법[12][13]은 컬러 히스토그램, 컬러 양자화(quantization), 컬러 비트 축소(color bit reduction) 기법들을 이용하여 문자 영역을 추출한다. 이러한 방법들은 추출될 컬러의 사전 지식(문자의 색상, 영역 비율 등)이 필요로 한다는 문제점이 있다. 그 이외에도 비디오에 존재하는 3차원 문자를 추출하기 위한 방법[14]도 제안되기도 하였다.

기존 문자 영역 추출 방법들의 문제점은 추출하고자 하는 문자의 사전 지식(문자의 크기, 색상, 서체, 등)이 필요하다는 것이다. 또한 많은 휴우리스틱 변수와 추출하고자 하는 문자의 사전 지식에 의해 임계치 값을 설정함으로써 알고리즘을 일반화하기 어렵다는 문제점이 있다. 본 논문에서는 문자의 지형학 특징점 추출 방법과 점-선-면 확장법을 이용하여 문자 영역을 추출함으로써 기존 문자 영역 추출의 문제점인 휴우리스틱 변수를 최소화하고 임계치 값을 일반화함으로써 일반화된 문자 영역 추출 방법을 제안 하고자 한다.

## 2. 문자 영역의 특성 분석

기존 논문에서 문자 영역 추출 방법의 일반화가 어려웠던 이유는 추출하고자 하는 문자의 색상, 서체, 크기 등과 같은 문자의 사전 지식 정보를 반영하여 알고리즘을 구성하기 때문에 구성된 알고리즘과 다른 서체, 색상의 문자 영역을 추출하기 위해서는 사용된 변수의 변경과 알고리즘 수정이 불가피하다. 따라서 문자 영역 추출 방법을 일반화한다는 것은 어렵다고 볼 수 있다. 본 논문에서는 문자의 지형학적 문자 특성을 다단계(Multi-level)로 추출하고 상호 관련된 변수를 일반화함으로써 문자 영역 추출 알고리즘을 일반화하는 방법을 제안하겠다.

문자 영역은 개별 문자의 구성원으로 이루어진 집합체로 볼 수 있다. 개별 문자의 명암도 특성은 그림1과 같이 배경과 문자의 명암도 차( $\alpha$ ), 문자 획의 폭( $\beta$ ) 그리고 문자간 거리( $\gamma$ )와 같은 3가지 특성 변수로 간략히 정의 할 수 있다. 여기서  $\alpha$ 값은 문자의 색상과 관련이 있고  $\beta$ 값은 문자의 서체와 관련이 있다고 볼 수 있다.

일반화된 문자 영역 추출을위해 앞에서 언급한  $\alpha$ ,  $\beta$ ,  $\gamma$  값을 일반화하는 방법을 모색하여야 한다. 일반화된  $\alpha$ ,  $\beta$ ,  $\gamma$  값을 구한다는 것은 어느 상황에도 적용 가능한  $\alpha$ -임계값(Threshold value),  $\beta$ -임계값,  $\gamma$ -임계값을 구한다는 이야기와도 같다. 다시 말해서 어느 문자에 대해서도  $\alpha$ -임계값,  $\beta$ -임계값,  $\gamma$ -임계값이 모두 만족할 수 있는 경우를 만들겠다는 것이다.

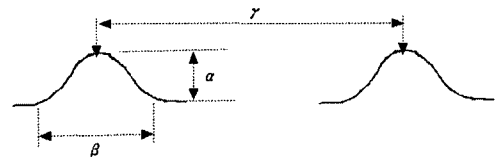


그림 1 문자의 지형학적 특성과 상호 관련된 변수의 특징

$\alpha$ -임계값과  $\beta$ -임계값은 문자의 색상과 서체에 따라서 동일 영상에서 다양하게 공존할 수 있다. 예를 들어, 그림2와 같이 배경이 복잡하고 문자 색상이 여러 종류가 혼합되어있을 경우 하나의  $\alpha$ -임계값 결정은 문제가 발생한다. 왜냐하면 임계값을 b로 결정할 경우 명암도가 b보다 낮은 문자는 포함되지 않을 수 있고, 만약 a보다 작게 임계값을 정할 경우 모든 문자 영역을 포함시킬 수 있으나 불필요한 배경 부분이 문자 영역으로 참여됨으로서 문제의 복잡도를 증가시킬 수 있다.

그러면 이와 같은 다양한 문자의 특성을 만족하기 위한  $\alpha$ -임계값,  $\beta$ -임계값,  $\gamma$ -임계값을 어떻게 구할 것인가가 중요한 문제로 등장한다. 본 논문에서는 기존의 지형학적 문자 특징점 추출 방법[15]을 수정 보완하여  $\alpha$ -임계값,  $\beta$ -임계값을 결정하는 방법을 사용하였다.

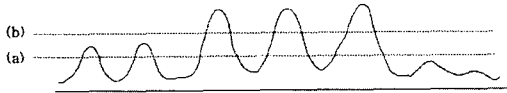


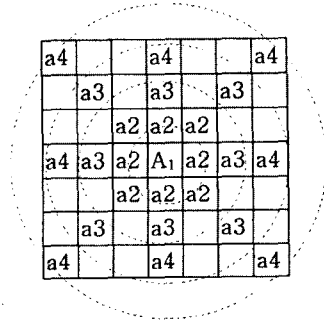
그림 2 문자의 색상과  $\alpha$ -임계값과의 관계

문자의 지형학적 특징점 추출 방법은 그림3(a)와 같이 마스크 형태로 표현할 수 있다. 이 마스크에  $\alpha$ ,  $\beta$ 에 관계된 정보를 관련시키면  $\alpha$ -임계값,  $\beta$ -임계값을 구할 수 있을 것으로 본다. 문자의 지형학적 특징을 보면 그림3(b)와 같이 배경 부분에서 문자 부분으로 단조 증가하다가 정점을 기점으로 감소함을 볼 수 있다. 따라서 마스크의  $A_1(=a_1)$  부분이 문자의 정점 부분에 있다고 가정하면 이웃 화소의 평균값  $A_2(3*3$ 마스크,  $a_2$ 의 평균값),  $A_3(5*5$ 마스크,  $a_3$ 의 평균값),  $A_4(7*7$ 마스크,  $a_4$ 의 평균값),... 은 점점 감소함을 볼 수 있을 것이다. 이것을 식으로 표현하면 식(1)과 같다. 이와 같은 기존의 지형학적 특징 추출 방법에 이웃 화소의 평균값 차등화 방법을 추가하여 차등 값  $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n$ 을 식(1)에 반영하면 식(2)와 같이된다.  $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n$ 을 일일이 결정한다는 것은 매우 어려운 문제이므로 모든 변수에 대하여 식(3)과 같이 임의의 값  $\alpha_{th} (= \alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n)$ 를 동일하게 설정하고  $\alpha_{th}$ 을 일정한 크기로 증가시키면서 문자의 색상에 대한 특성을 반영한다.  $\alpha_{th}$ 값과 관련된 사항은 3장 3.1절에서 자세히 설명하겠다.

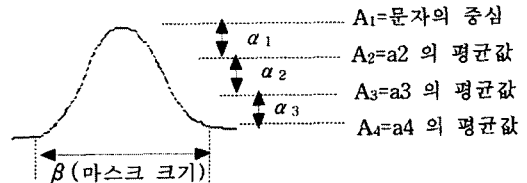
$\beta$ -임계값은 문자의 서체와 관련이 있다고 했다. 즉, 문자의 획 폭과 관련이 있다. 따라서 그림3(a)의 마스크 크기를 증가시키면 문자의 획 폭에 대한 문제를 해결할 수 있다. 그러면 마스크의 크기가 계속 커진다면 문제는 없는가?. 여기에 대한 문제점은 문제의 복잡성이 증가하고 마스크 연산에 대한 처리 시간이 너무 많이 소요되는 문제점이 발생한다. 긴 처리 시간 문제는 그림 3(a)와 같이 표시된 부분만을 계산에 참여시킴으로써 처리 시간을 단축한다. 마스크의 크기에 관련된 사항은 4장 4.1절에서 자세히 설명하겠다.

$\gamma$ -임계값은 문자의 크기와도 관련이 있다고 볼 수 있으나 문자 획과 문자 획간의 거리를 말하는 것으로  $\gamma$ -임계값을 결정할 때 모든 문자에 대하여 처리가 가

능하도록 보통 넓은 범위의  $\gamma$ -임계값으로 정한다.



(a) 마스크 표현 형태



(b) 문자의 지형학적 표현 형태

그림 3 문자의 특징 마스크와  $\alpha$ ,  $\beta$  값과의 관계

$$\text{Feature-Point} = \begin{cases} 1 : \text{if } (A_1 > A_2 > A_3 \dots A_n) \\ 0 : \text{otherwise} \end{cases} \quad (1)$$

$$\text{Feature-Point} = \begin{cases} 1 : \text{if } ((A_1 > A_2 + \alpha_1) \text{ and } (A_2 > A_3 + \alpha_2) \text{ and } (A_3 > A_4 + \alpha_3) \dots (A_n > A_{n+1} + \alpha_n)) \\ 0 : \text{otherwise} \end{cases} \quad (2)$$

$$\text{Feature-Point} = \begin{cases} 1 : \text{if } ((A_1 > A_2 + \alpha_{th}) \text{ and } (A_2 > A_3 + \alpha_{th}) \text{ and } (A_3 > A_4 + \alpha_{th}) \dots (A_n > A_{n+1} + \alpha_{th})) \\ 0 : \text{otherwise} \end{cases} \quad (3)$$

$\alpha_{th} = \alpha_{th+5}$   
 $(0 < \alpha_{th} < \alpha_{max})$

### 3. 일반화된 문자 및 비디오 자막 영역 추출 방법

일반화된 문자 및 자막 영역 추출 방법은 그림4와 같이 크게 2단계로 분류 할 수 있다. 문자의 지형학적 특징점을 이용한 문자 후보 영역 추출 단계와 추출된 후보 문자 영역을 검증하는 문자의 후보 영역 검증 단계로 분류 할 수 있다.

문자 후보 영역 추출 단계는 다음과 같이 3단계로 구성되어 있다. 문자의 지형학적 특징점 추출 단계, 점-선-면 확장 방법을 이용한 레벨별 문자 영역 추출 단계,

각 레벨에서 추출된 문자 영역 중 중복된 영역을 하나의 영역 합병하는 문자 후보 영역 추출 단계로 구성되어 있다. 문자의 지형학적 특징점 추출 단계는 식(3) 조건을 만족하는 화소에 대하여 레벨( $\alpha$ -th 값)별로 지형학적 특징 점을 추출한다. 추출된 문자의 지형학적 특징 점을 점-선-면 확장법을 이용하여  $\gamma$ -임계값 보다 큰 지형학적 특징점들 상호 연결하고, 형성된 선 영역을 면 영역으로 확장하여 레벨별 면 영역을 구한다. 각 레벨별로 추출된 면 영역들 중 중복된 면 영역을 하나로 합병하여 문자의 후보 영역을 추출한다.

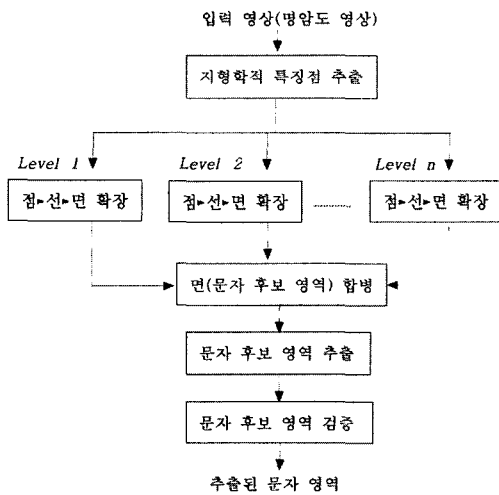


그림 4 문자 영역 추출 전체 흐름도

추출된 문자 후보 영역은 대단위 영역화 검증 단계를 통하여 각 영역이 문자 영역으로서 적합한지를 조사하여 최종적으로 문자 영역을 추출한다. 지금까지 설명한 내용에 대하여 3.1절 3.2절에서 일반화된 변수를 사용한 문자 영역 추출 방법 설명하겠다.

### 3.1 문자의 후보 영역 추출 방법

문자의 후보 영역 추출 알고리즘은 2단계로 구성되어 있다. 문자의 지형학적 특징점을 추출하는 단계1과, 점-선-면 확장 방법을 이용하여 문자의 후보 영역을 추출하는 단계2로 구성된다.

단계1과 관련된 임계값 변수는  $Lvalue$ (Level value),  $Ms$ (Mask Size),  $Interval$ 이며 이 변수는  $\alpha$ -임계값,  $\beta$ -임계값,  $\gamma$ -임계값과의 상호 관련이 있다.

$Lvalue$  변수는  $\alpha$ -임계값과 관련이 있으며 문자의 정점과 배경 사이의 명암차 임계값을 어느 정도까지 반영할 것인가에 대한 값이다.  $A_1, A_2, A_3, \dots, A_n$ 에 대한 명암

차를 반영하기 위한 값을  $\alpha$ -th로 정하고 이 값을 일정한 크기로 증가시키면서  $Lvalue$ 까지 문자의 명암차의 특성을 반영한다. 특징점의 중복 추출과 계산량을 줄이기 위하여  $\alpha$ -th 을 5씩 증가하여 사용하였다.

$Ms$  변수는  $\beta$ -임계값과 관련이 있으며 어느 정도의 문자 획을 처리할 것인가에 대한 값을 나타낸다.  $Ms$  크기에 따라  $A_n$ 값이 정해진다. 예를 들어, 마스크 크기가  $5 \times 5$ 이면  $Ms$ 는 5가 되며  $n$ 값은 식 $[Ms/2+0.5]$ 에 의하여 3이 된다.

$Interval$  변수는 문자의 탐색 범위를 나타내며 최소한 2번은 문자 부분을 탐색한다는 기본 원칙을 가지고 결정한다. 왜냐하면 영상의 불규칙한 특성상 문자 부분에서 후보 영역 추출의 실패를 방지하기 위해서 문자 부분에 대하여 2번은 탐색하도록 결정한다. 이렇게 일정한 간격으로 탐색하는 것을 라인 샘플링 방법(line sampling) 이라고 한다. 단계1은 그림 5에서 플로우 차트(flow chart) 형식으로 자세히 설명되고 있다

// 단계 1 : 특징 추출 단계 : 문자의 지형학적 특징 정보 추출

$Lvalue = ?$  ; //  $\alpha$ -임계값 관련

$Ms = ?$  ; //  $\beta$ -임계값 관련

$Interval = ?$  ; //  $\gamma$ -임계값 관련

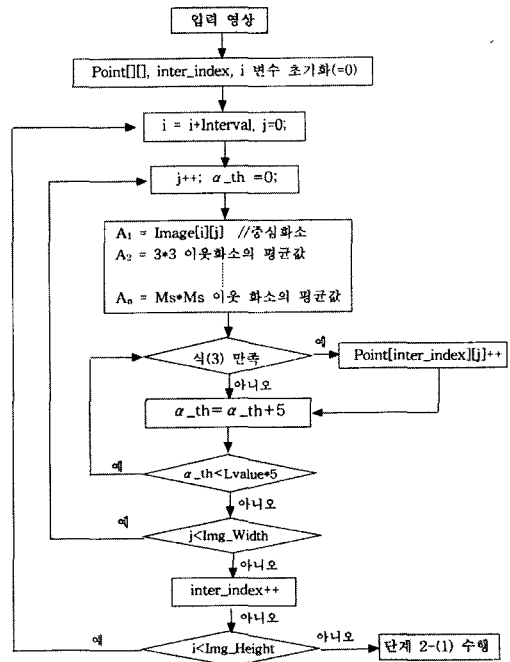


그림 5 문자의 지형학적 특징 정보 추출 (단계 1)

단계2는 각 레벨별로 문자의 영역을 추출하는 단계이며, 점-선-면 확장법과 각 레벨별로 추출된 면 영역의 합병 방법을 이용하여 문자의 후보 영역을 추출한다. 후보 문자 영역 추출과 관련된 임계값 변수는 *Point\_inter\_th*와 *Block\_merge\_th*이다.

*Point\_inter\_th* 값은  $\gamma$ -임계값과 관계가 있으며 문자 획과 획 사이의 거리에 대한 임계값을 나타낸 것이다. 사전 지식이 없을 경우에는 모든 범위를 포함할 수 있도록 범위를 크게 정한다.

*Block\_merge\_th* 값은 각 레벨로 형성된 문자의 영역들이 어느 정도 중첩되면 같은 하나의 영역으로 합병할 것인가에 대한 임계값이다. 단계2는 그림6, 7에서 자세히 설명되고 있다

각 단계별로 사용되는 임계값 변수는 4.1절에서 표로 정리하여 상세히 설명하겠다.

// 단계 2-(1) : 자막 후보 영역 추출 단계 : 점-선-면 방법 이용  
*Point\_inter\_th* = ?; //  $\gamma$ -임계값 관련

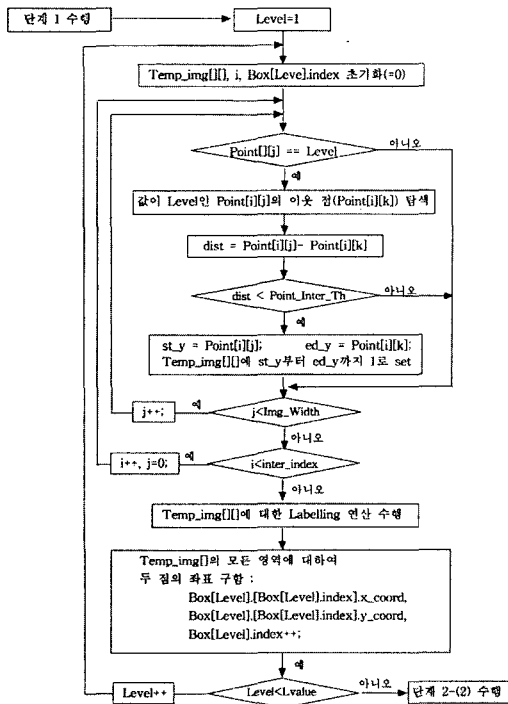


그림 6 자막 후보 영역 추출 (단계 2-(1))

// 단계 2-(2) : 자막 후보 영역 추출 단계- 각 블록 합병 및

블록 검증

*Block\_merge\_th* = ?; // 중첩된 면의 합병조건 임계값

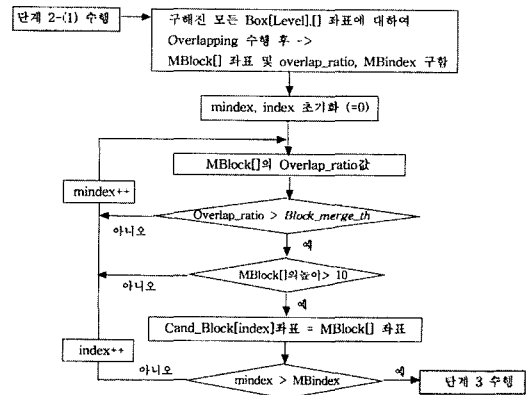


그림 7 각 블록 합병 및 블록 검증 (단계 2-(2))

문자가 포함된 원 영상은 그림8에서 보여주고 있으며 라인 샘플링 기법(점선 부분의 행만 탐색)을 이용하여 일정한 간격으로 행을 선택하여 처리 범위를 축소한다. 선택된 각 행에서 단계1(지형학적 특징추출)을 수행한 결과 그림9과 같이 지형학적 특징점이 추출됨을 볼 수 있으며 Reduce[=영상 높이/Interval] 만큼 축소를 볼 수 있다. 그림10은 단계2-1(점-선-면 확장)중에서 점에서 선으로 형성시킨 결과를 보여 주고있으며, 선의 연결하여 형성된 면의 결과는 그림11에서 보여주고 있다. 이렇게 각 레벨에서 추출된 면은 단계2-2(면의 합병)를 수행하여 하나의 통합된 문자 후보 영역을 추출하며, 그 결과는 그림 12에서 보여주고 있다.

3.2 추출된 후보 영역 검증

추출된 문자의 후보영역은 검증을 통하여 최종적으로 문자 영역으로 판단한다. 문자 영역 검증 시 사용된 변



그림 8 원 영상 및 라인 샘플링 기법

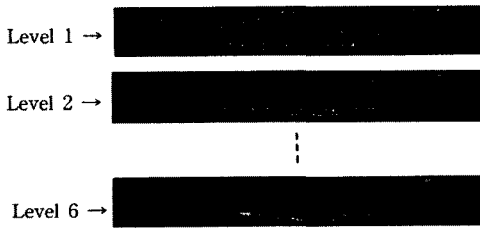


그림 9 추출된 지형학적 문자의 특징점

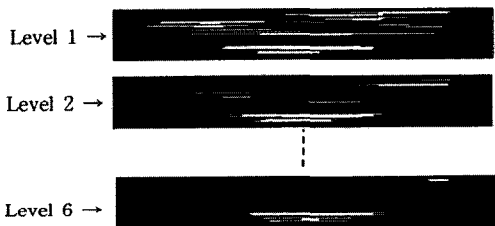


그림 10 각 레벨별로 선 형성

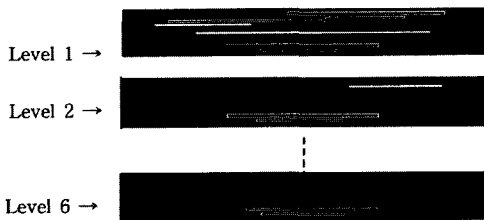


그림 11 각 레벨별로 형성된 면 영역 추출

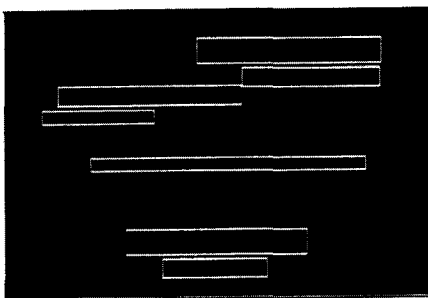


그림 12 면 영역 합병 결과

수는 *Block\_thresh\_th*, *Area\_min\_th*, *Area\_max\_th*, *BW\_ratio\_th*, *Verti\_overlap\_th*, *Char\_Number\_th*이며, 검증 단계는 2단계로 구성된다.

첫 번째 단계는 추출된 후보 영역에 대하여 이진화하고 라벨링 연산을 수행하여 문자의 구성 요소를 구하는 단계이다. 이진화시 필요한 임계값 변수는 *Block\_Thresh\_th* 변수이며 이진화시 문자 영역 대 배경 영역을 어느 정도로 이진화 할 것 인가를 정해주는 임계값은 이다. 이진화 후 라벨링 연산을 통하여 구해진 문자의 구성 요소를 추출한다. 추출된 문자의 구성요소에 대하여 *Area\_min\_th*, *Area\_max\_th*, *BW\_ratio\_th* 임계값을 적용하여 문자의 구성 요소로 적합한지 조사하고 적합치 않은 구성 요소는 제거한다.

두 번째 단계는 문자의 구성 요소에 대하여 대단위 영역화 방법을 통하여 대단위 문자 행을 구하고 최종적으로 문자 행에 포함된 문자의 개수를 비교하여 문자 영역을 추출한다. 대단위 영역화에 사용되는 변수는 *Verti\_overlap\_th*이며 구성 요소의 수직 부분의 중첩 비율에 대한 임계값이다. 대단위 영역화 방법은 구성 요소의 수직 부분이 중첩 비율이 *Verti\_overlap\_th* 보다 크면 같은 행의 문자 구성 요소로 보고 대단위 영역에 소속시킨다. 중첩 비율이 너무 작으면 다른 대단위 영역을 생성한다.

대단위 영역화 방법에서 구해진 대단위 영역 안에서 문자의 개수가 *Char\_Number\_th* 임계값 보다 크면 그 대단위 영역을 문자 영역으로 간주하고 문자 영역으로 추출한다.

```
// 단계 3 : 문자의 후보 영역 검증 단계
Block_thresh_th = ?; // 이치화 임계값
Area_min_th = ?; // 최소 면 영역 임계값
Area_max_th = ?; // 최대 면 영역 임계값
BW_ratio_th = ?; // 면 영역의 문자 대 비 문자의 비 임계값
Verti_overlap_th = ?; // 대단위 영역화의 수직 영역 중첩 비율
Char_Number_th = ? // 문자 영역으로 볼 수 있는 문자의 개수
```

```
Verification_Processing()
{
    // 이치화 및 라벨링 연산을 통해 문자 자소 단위 영역 추출
    Char_Component[CCinx].xy_corr
    =Threshold_and_Labelling_Operation(Cand_Block,
    Block_Thresh_th);
```

```
// 전처리 : 자소 단위 영역이 조건에 맞지 않으면 제거
for ck =0 to CCinx // 추출된 문자 단위의 자소 총 개수
```

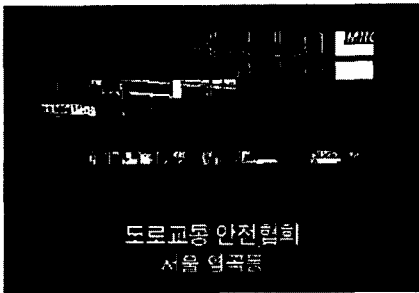
```

{
    area = Char_Component[ck];
    bw_ratio = char_area / area;
    if(area > Area_min_th && area < Area_max_th) ||
    (bw_ratio > BW_ratio_th)
        discard_current_component(ck);
}

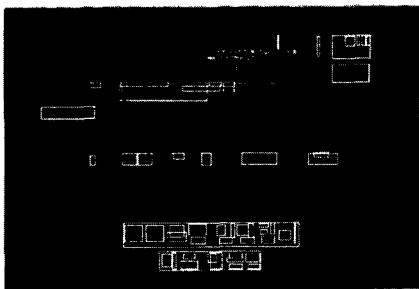
// 문자 자소 단위 영역 통합에 의한 문자 단위 영역 형성
for ck = 0 to CCinx
    Large_Char_Group[LGinx].num
    =Construction_Character_Group(Char_Component,
    Verti_overlap_th);

// 추출된 문자 단위 영역의 개수 조사 -> 문자 영역 여부 판단
tx = 0;
for all lc(=0~LGinx)
    if(Large_Char_Group[lc].num >
        Char_Number_th)
        Text_region[tx++] = Large_Char_Group[lc];
}
    
```

추출된 문자의 후보 영역에 대하여 이진화를 수행한 결과는 그림13(a)에서 보여 주고 있다. 라벨링 연산을 통하여 구해진 문자의 구성 요소에 대하여 대단위 영역화 방법을 수행한 결과는 그림13(b)에서 보여 주고 있다.



(a) 이진화된 결과



(b) 문자의 구성 요소 추출 및 대단위 영역화

그림 13 추출된 문자의 후보 영역 검증



그림 14 추출된 문자의 후보 영역

각 대단위 영역에 포함된 문자의 개수를 조사하여 미리 설정된 임계값  $Char\_Number\_th$  보다 크면 그 대단위 영역을 문자 영역으로 간주한다. 대단위 영역에 포함된 문자의 개수를 검증 후 추출된 문자 영역은 그림14에서 보여 주고 있다.

#### 4. 실험 및 고찰

##### 4.1 사용된 변수의 임계값 분석 및 정리

본 논문에서 사용된 임계값 변수는 표1과 같이 11개이며 사용자가 추출하고자 하는 자막에 대한 사전 지식을 알고 있을 경우와 모를 경우로 구분하여 설명하겠다. 사용된 임계값 변수를 단계별로 자세히 설명하면 다음과 같다.

문자의 후보 영역 추출 과정에 사용된 변수는  $Lvalue$ ,  $Ms$ ,  $Interval$ ,  $Point\_inter\_Th$ ,  $Block\_merge\_th$ 이다.

$Lvalue$ (=Level value) 임계값 변수의 경우 사전 지식이 없을 경우  $Lvalue$  값을 0~255의 값을 정할 수가 있다. 이렇게 되면 계산 량이 많아지고 중복된 특징점들이 추출될 수 있으므로 본 논문에서는  $Lvalue$ 에 5를 곱한 값(= $Lvalue*5$ )을 사용하였다. 사전 지식이 없을 경우  $Lvalue$ 는 6정도로 정해도 충분한 값으로 실험 결과 볼 수 있었다. 즉, 각 이웃 화소의 평균값을 최대 30(= $6*5$ )까지 하겠다는 것이다.  $Lvalue$ 에 대한 사전 지식을 갖고있을 경우에는 사용자 정의값을 사용한다.

$Ms$  임계값 변수는 마스크 크기와 관련이 있으며, 실험결과 마스크 크기는 획의 폭 1.5~2배 정도 처리할 수 있음을 알 수 있었다. 마스크의 크기가 4일 경우 문자 획의 폭은 6~8화소 정도가 되더라도 처리가 가능함을 볼 수 있었다. 따라서 3\*3\*9\*9 마스크 크기이면 문자 획의 폭에 대한 문제를 해결 할 수 있음을 볼 수 있었다.  $Ms$ 에 대한 사전 지식이 없을 경우  $Ms$  값을 최대

9로 정하여 사용하였고, 사전 지식이 있을 경우에는 사용자 정의 값으로 사용한다.

*Interval* 임계값 변수는 탐색 범위를 어느 정도 줄일 것인가와 관련이 있다. 영상에서 최소 문자 판별 단위를 10 화소로 정할 경우 최소한 2번은 탐색한다는 원칙 하에서 문자에 대한 사전 지식이 없을 경우 5로 결정한다. 사전 지식이 있을 경우에는  $5 < \text{지정} < \text{Height} * 0.05$  범위 사이에서 정한다.

*Point\_inter\_th* 임계값 변수는 문자 획과 획 사이의 거리에 대한 임계값으로 문자에 대한 사전 지식이 없을 경우 최대  $\text{Width} * 0.3$  값으로 결정한다. 왜냐하면 모든 연결 가능한 지형학적 특징 점들을 연결하여 선으로 형성시킴으로서 문자일 가능성 있는 모든 문자를 모두 포함시키겠다는 생각이다. 사전 지식이 있을 경우에는 사용자 지정 값으로 정하여 사용한다.

*Block\_merge\_th* 임계값 변수는 각 레벨에서 구해진 후보 영역을 합병할 때 사용하는 변수로 사전 지식이 없을 경우 실험 결과 80%가 적정함을 볼 수 있었으며, 사전 지식이 있을 경우 70~90%사이의 값을 지정한다.

추출된 후보영역에 대한 검증과 관련된 변수는 *Block\_thresh\_th*, *Area\_min\_th*, *Area\_max\_th*, *BW\_ratio\_th*, *Verti\_overlap\_th*, *Char\_Number\_th*이다.

*Block\_Thresh\_th* 임계값 변수는 이진화와 관련이 있

으며 사전 지식이 없을 경우 이진화 결과가 35% 정도 문자 영역이 점유할 때 적정히 이진화 됨을 볼 수 있었다. 사전 지식이 있을 경우에는 사용자가 지정하는 임계값을 사용한다.

*Area\_min\_th*, *Area\_max\_th* 임계값 변수는 추출된 문자의 구성 요소의 최대 최소 범위를 정하여 너무 작거나 너무 큰 경우에는 구성요소에서 제거하는 전처리 과정에서 필요한 임계값이다. 사전 지식이 없을 경우에는  $\text{Total\_area} * 0.01 \sim \text{Total\_area} * 0.1$  범위로 정하고 이 이상 벗어나는 구성 요소는 제거한다. 사전 지식이 있을 경우에는 사용자 지정값으로 정한다.

*BW\_ratio\_th* 임계값 변수 역시 구성 요소의 전처리를 위한 임계값이다. 사전 지식이 없을 경우에 구성요소에 이진화된 문자 영역의 비율이 50~70% 정도의 비율이 적정함을 볼 수 있으며 이 범위를 벗어나는 영역은 제거한다. 사전 지식이 있을 경우에는 사용자 지정값을 사용한다.

*Verti\_overlap\_th* 임계값 변수는 대단위 영역화 방법에 필요한 임계값으로 사전 지식을 없을 경우 실험 결과 수직 중첩율이 70% 정도가 적당함을 알 수 있었다. 사전 지식이 있을 경우에는 사용자 지정값을 사용한다.

*Char\_Number\_th* 임계값 변수는 대단위 영역화에 포함된 문자가 몇 개까지 포함되면 문자 영역으로 볼 것인가에 대한 임계값으로 사전 지식이 없을 경우에는 3

표 1 사용자 변수와 관계

| 사용 변수       |                                       | 문자에 대한 사전 지식   |  | 수행단계  |
|-------------|---------------------------------------|--|--|-------|
| 구분          | 변수                                    | 모름경우(기본값)  | 알 경우(지정값)                              | 단계    |
| 후보 영역 추출 단계 | <i>Lvalue</i>                         | 6  | $0 < \text{지정} < 6$                    | 1     |
|             | <i>Ms</i>                             | 9  | $3 < \text{지정} < 9$                    | 1     |
|             | <i>Interval</i>                       | 5  | $5 < \text{지정} < \text{Height} * 0.05$ | 2     |
|             | <i>Point_inter_th</i>                 | $\text{Width} * 0.3$                                   | 지정값                                    | 2-(1) |
|             | <i>Block_merge_th</i>                 | 80%  | $70\% < \text{지정} < 90\%$              | 2-(2) |
| 후보 영역 검증 단계 | <i>Block_thresh_th</i>                | 35%  | 지정값                                    | 3     |
|             | <i>Area_min_th</i> <i>Area_max_th</i> | $\text{Total\_area} * 0.01$ $\text{Total\_area} * 0.1$ | 지정값                                    | 3     |
|             | <i>BW_ratio_th</i>                    | 50~70%   | 지정값                                    | 3     |
|             | <i>Verti_overlap_th</i>               | 70%  | 지정값                                    | 3     |
|             | <i>Char_Number_th</i>                 | 3  | $3 \leq \text{지정}$                     | 3     |

where :: *Width* : 영상의 가로 크기  
*Height* : 영상의 세로 크기  
*Total\_area* :  $\text{Width} * \text{Height}$



개, 사전 지식을 갖고 있을 경우에는 3개 이상으로 정하여 최종 문자 영역으로 판단한다.

**4.2 실험 환경 및 결과**

실험 환경은 Windows NT상에서 550 MHz 속도를 갖는 SGI Pentium PC로 구현하였다. Visual C++6.0 언어를 사용하였으며 비 압축 비디오 영상을 실험 영상으로 사용하였다. 비디오 영상은 RT5 Encoding 보드를 사용하여 file.mpg 파일을 획득한 후 소프트웨어적으로 Decoding하여 실험에 사용하였다. 본 논문에서 제안된 방법이 프레임 차를 이용한 방법이 아니며 비디오 자막의 경우 최소 1초 이상은 화면상에 유지되기 때문에 1초당 1프레임 간격으로 영상을 획득했다.

실험에 사용한 비디오 영상은 뉴스 2,334 프레임, 영화 4,221 프레임, 다규멘터리 3,230 프레임, 스포츠 2,450 프레임에 대하여 실험하고 그 결과를 표2에서 보여 주고 있다. 본 논문의 객관성을 살펴보기 위하여 문자 영역 추출율과 처리시간에 대하여 비교 실험하였다.

처리 시간면에서 비교해 보면, 사전 지식이 없는 경우 후보 영역의 추출 개수가 사전 지식이 없는 경우 보다 많게 되기 때문에 검증에 소요되는 시간이 많아짐을 볼 수 있다.

영역 추출율면에서 비교해 보면, 사전 지식이 없는 경우와 사전 지식이 있는 경우가 큰 차이를 보이지 않음을 볼 수 있다. 그 이유는 사전 지식이 없는 경우는 단지 후보 영역을 조금 더 찾아 준다는 차이점밖에 없기 때문에 이 문제는 대부분 검증 단계에서 처리함으로써 영역 추출율에서 큰 차이를 보이지 않음을 볼 수 있다.

4종류의 비디오 영상 대하여 실험한 결과 문자의 사전 지식 정보와 관계없이 후보 문자 영역 추출율은 100%, 검증을 통한 최종 문자 영역 추출율은 98% 이상임을 볼 수 있었다.

표 2 제안된 연구 방법의 실험 결과

|              |           | 뉴스 영상 | 영화 영상  | 다규멘터리  | 스포츠 영상 |
|--------------|-----------|-------|--------|--------|--------|
| 사전 지식이 없는 경우 | 자막 영역 추출율 | 99.0% | 98.3 % | 98.4 % | 99.1%  |
|              | 처리 시간     | 0.29초 | 0.32초  | 0.38초  | 0.23초  |
| 사전 지식이 있는 경우 | 자막 영역 추출율 | 99.2% | 98.6%  | 98.8%  | 99.2%  |
|              | 처리 시간     | 0.14초 | 0.17초  | 0.15초  | 0.13초  |

제안된 방법을 기존 방법과 비교 분석하기 위하여 특성이 다른 3개의 기존 연구 방법과 비교 분석한 결과들

보면 표3과 같다. 기존의 연구 방법은 전체 영상을 처리함으로써 처리 시간이 많이 소요될 뿐 아니라 사전 지식에 대한 의존도가 매우 높아 알고리즘을 일반화하여 구현하기가 쉽지 않은 단점이 있다.

본 논문에서 제안한 방법의 경우 부분 영상을 처리하고 추출하고자 하는 문자의 사전 지식 없이도 문자 영역을 추출함을 볼 수 있었으며, 만약 문자에 대한 사전 지식을 사용할 경우 처리 시간이 짧아짐을 볼 수 있었다.

표 3 기존의 방법과 비교 분석

|           | 제안된 방법       | Lienhart 방법 <sup>[10]</sup> | Kurakake 방법 <sup>[11]</sup> | Kim 방법 <sup>[13]</sup> |
|-----------|--------------|-----------------------------|-----------------------------|------------------------|
| 처리 방법     | 다단계 특정 추출 방법 | 영역 합병 및 분할 방법               | 프레임 차 방법                    | 컬러축소 방법                |
| 처리 범위     | 부분영상         | 전체 영상                       | 모든 프레임 + 전체 영상              | 전체 영상                  |
| 사전 지식 정보  | 필요없음         | 필요함                         | 필요함                         | 일부 필요함                 |
| 변수의 일반화   | 가능           | 불가능                         | 불가능                         | 일부 가능                  |
| 사용된 변수 수  | 11           | 13±1                        | 14 ±1                       | 16±1                   |
| 알고리즘의 일반성 | 좋음           | 보통                          | 나쁨                          | 보통                     |

**5. 결론**

본 논문의 경우 추출하고자 하는 문자에 대한 사전 지식을 모를 경우에도 문자 영역을 추출함을 볼 수 있었으며, 사용된 변수의 임계값을 일반화함으로써 문자 영역 추출 방법의 일반화를 이룰 수 있었다.

향후 연구 방향은 두 가지로 진행하고자 한다. 첫째, 현재 명암도 영상에서 추출하고 있는 지형학적 특징 추출 방법을 컬러 영상에서 추출하는 방법에 대하여 연구를 진행하고자 한다. 둘째, 각 레벨에서 추출된 블록 합병 방법을 최적화 하는 방법에 대하여 연구하고자 한다.

**참고 문헌**

[1] HongJiang Zhang, Shuang Yeo Tan, Stephen W. Smoliar and Gong Yihong, "Automatic parsing and indexing of news video," Multimedia System, Vol.2, pp.256-266, 1995.

[2] HongJiang Zhang, C.Y.Low, S.W.Smoliar and J.H.Wu, "Video Parsing, Retrieval and Browsing : An Integrated and Content-based Solution," Proc. ACM Multimedia 95, San Francisco, CA, pp.15-24, Nov. 1995.

- [3] Michael A. Smith, and Takeo Kanade, "Video Skimming for Quick Browsing based on Audio and Image Characterization," Carnegie Mellon Univ., Technical Report CMU-CS-95-186, July 1995.
- [4] Jun Ohya, Akio Shio and Shigeru Akamatsu, "Recognizing Characters in Scene Images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.16, No.2, pp.214-220, 1994.
- [5] B. Yu and A. Jain, "Robust and fast skew detection algorithm for generic documentation," Pattern Recognition, Vol.29, pp.1599-1629, 1996.
- [6] B. Yu and A. Jain and M. Mohiuddin, "Address block location on complex mail pieces," Proc. of the 4th Int. Conf. on Document Analysis and Recognition, Ulm, 1997.
- [7] A. Jain and S. Bhattacharjee, "Text segmentation using Gabor filters for automatic document processing," Machine Vision Application, Vol.5, pp.169-184, 1992.
- [8] I. Pitas and C. Kotropoulos, "A textured-based approach to the segmentation of semitic image," Pattern Recognition. Vol.25, pp.929-945, 1992.
- [9] Y. Zhong, K. Karu, and A. Jain, "Locating text in complex color images," Pattern Recognition, vol.28, pp.1523-1535, 1995.
- [10] Rainer Lienhart and Frank Stuber, "Automatic text recognition in digital videos," Proc. of the SPIE, Image and Video Processing IV, Vol. SPIE2666, pp.180-188, San Jose, 1996.
- [11] Shoji Kurakake, Hidetaka Kuwano and Kazumi Odaka, "Recognition and visual feature matching of text region in video for conceptual indexing," Proc. of the SPIE, Storage and Retrieval for Image and Video Database V, vol. SPIE3022, pp.368-378, San Jose, 1997.
- [12] Yu Zhong, Kalle Karu and Anil K. Jain, "Location text in complex color images," Pattern Recognition, Vol.28, No.10, pp.1523-1535, 1995.
- [13] Hae-Kwang Kim, "Efficient automatic text location method and content-based indexing and structuring of video database," Journal of Visual Communication and Image Representation, Vol.7, No.4, pp.336-344, Dec. 1996.
- [14] Jun Ohya, Akio Shio and Shigeru Akamatsu, "Recognizing characters in scene images," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.16, No.2, pp.214-220, 1994.
- [15] Byung Tae Chun, Younglae Bae and Tai-Yun Kim, "Text Extraction in Videos using Topographical Features of Characters," The 8th IEEE Int. Conf. on Fuzzy System(Fuzz-IEEE'99), Vol.2, pp.1126-1130, Seoul, 1999.



전 병 태

1986년 2월 한남대학교 전산과(학사).  
 1989년 8월 숭실대학교 전산과(석사).  
 1997년 3월 ~ 현재 고려대학교 컴퓨터  
 학과(박사과정). 1992년 5월 IR52 장영  
 실 상 수상(과기부 장관상). 1989년 9월  
 ~ 현재 한국전자통신연구원 컴퓨터·소  
 프트웨어 기술연구소 선임연구원 재직중. 관심분야는 동영  
 상 처리, 컴퓨터 비전, 멀티미디어 영상처리 등



배 영 래

1976년 서울대학교 해양물리전공(이학  
 사). 1986년 한양대학교 전자계산 전공  
 (공학석사). 1992년 Univ. of Kent 전자  
 공학 전공(공학석사). 1995년 Univ. of  
 Kent 전자공학 전공(공학박사). 1976  
 년~1979년 해군장교. 1979년 ~ 1980년  
 삼성. 1980년 KIST입소, 현재 ETRI 컴·소 연구소 영상  
 정보처리연구팀장. 1997년 배재대학교 겸임교수. 관심분야  
 는 영상처리, 컴퓨터 비전, 병렬처리



김 태 윤

1981년 고려대학교 산업공학과 학사.  
 1983년 Wayne state University 전산  
 과학과 석사. 1987년 Auburn  
 University 전산과학과 박사. 1998년 ~  
 현재 고려대학교 컴퓨터학과 교수. 관심  
 분야는 전자상거래, 컴퓨터 네트워크,  
 EDI, 이동통신 등