

# 동영상에서 인물식별을 위한 얼굴검출 알고리즘 구현

조 미 남,<sup>†</sup> 지 유 강,<sup>‡</sup>  
동신대학교

## Implementation of a face detection algorithm for the identification of persons

Mi-nam Cho,<sup>†</sup> Yoo-kang Ji,<sup>‡</sup>  
Dongshin University

### 요 약

영화, 드라마와 같은 비디오에서 등장하는 인물을 검출하고 인식하는 기술은 일반 사용자를 대상으로 PVR(personal video recorder), 모바일 폰과 멀티미디어 공급자를 위한 얼굴 영상의 데이터베이스 관리와 같은 응용분야에 이용될 수 있다. 본 논문에서는 비디오 등장인물을 자동으로 검색하는 얼굴검출 기법을 제안한다.

본 논문에서 제안한 얼굴검출 기법은 크게 세 단계로 구성하였다. 첫 번째, 정지 영상을 생성한 후 장면전환점 검출단계, 두 번째, 색상정보를 이용한 얼굴검출 단계, 마지막으로 얼굴영역을 이용한 얼굴 특징점 검출단계로 구성하였다. 실험 결과 다양한 환경에서 성공적으로 얼굴을 검출하였으며, 기존의 색상정보를 이용한 얼굴 검출기법보다 성능이 개선된 것을 확인할 수 있었다.

### ABSTRACT

The technique, which is able to detect and recognize characters in videos such as a movie or TV drama, can be used for applications which are database management of a general user's facial images for the suppliers of PVR(personal video recorder), mobile phones, and multimedia, etc.

In this paper, we propose a face detection algorithm. It searches the character through cast indexing when the scene is changed in video. It is consisted of three stages. The first step is the detection-step of the scene change after producing a paused image. The second step is the face detection-step using color information. The final step is the detection-step which detects its features by the facial boundary. According to the experimental result, it has detected faces in different conditions successfully and more advanced than the existing other one that are using only color information.

**Keywords:** face detection, side-view face detection, scene change, facial feature, video

## 1. 서 론

동영상 압축 기술 및 통신기술, 대용량 저장매체 및 컴퓨터의 성능 향상, 카메라 기술의 발전으로 디지털 비디오 정보의 활용이 급속하게 성장하고 있다. 영상 처리를 이용한 얼굴 인식은 시각적인 내용 정보를 바

탕으로 서로 다른 사람을 구분한다. 이때, 얼굴의 차이를 보고 다른 사람이라고 인지한다. 대규모의 사진 데이터베이스나 동영상 등에서 사람 얼굴의 특징인 얼굴을 검출하고 인식하는 기술은 일반 사용자를 대상으로 한 PVR(personal video recorder), 모바일 폰과 멀티미디어 정보 검색을 효율적으로 검색하기 위한 비디오 관리 시스템에 이용될 수 있다.

지금까지 연구되었던 얼굴검출 기법은 얼굴의 위치, 크기, 방향, 빛의 조건들에 대한 영향을 고려하지

접수일(2010년 5월 18일), 게재확정일(2010년 10월 1일)

<sup>†</sup> 주저자, mncho1004@naver.com

<sup>‡</sup> 교신저자, neobacje@gmail.com

않고 단일한 얼굴의 위치를 구분하여 얼굴을 구성하는 눈, 코, 입 등의 요소들을 확인하는 것으로 얼굴을 검출하였다.

기존에 연구되었던 얼굴검출 방법으로는 특징 기반 방법들과 영상 기반 방법들로 구분될 수 있다. 특징 기반 방법으로는 에지와 그레이 영상에 대한 검출 방법으로서 템플릿 매칭 기법<sup>[1]</sup>, 신경망<sup>[2]</sup>, 학습에 기반을 둔 방법<sup>[3]</sup> 등이 있다. 그러나 이러한 방법은 특정한 크기와 정해진 시각 안의 눈, 코, 입등의 거리 및 위치 등을 이용하기 때문에 연산량이 많은 단점을 가지고 있다. 그리고 영상 기반 방법은 최근에 칼라 정보에 기반을 둔 얼굴 검출 방법이 제시되고 있다. 칼라 정보에 기반을 둔 얼굴검출 방법에서 얼굴 영역들은 피부색의 특성에 의해 분리될 수 있으며, 분리된 후보 영역과 다른 템플릿 또는 얼굴의 기하학적 지식을 이용할 수 있고, 또한 칼라 정보를 이용한 얼굴검출 방법은 얼굴의 특징을 빠르고 쉽게 얻을 수 있다. 다른 방법에 비해 연산량을 줄일 수 있어 실시간으로 구현될 수 있는 장점을 가지고 있다. 그러나 칼라정보를 이용한 방법은 조명의 변화에 민감하기 때문에 얼굴과 유사한 칼라 분포를 모두 검출해 낸다는 단점을 가지고 있다.

본 논문은 비압축된 비디오 칼라 영상에서 등장인물의 얼굴을 검출하기 위해 히스토그램 방식을 이용한 장면전환점을 검출하여 정지영상을 생성한 후 색상 정보 기반의 방법을 사용한다. 비디오 영상에서 장면전환점 검출로 정지영상을 생성하면 프레임 기반의 유사성을 컷 추출로 인해 상당수 배제할 수 있어 색상정보를 이용한 얼굴검출 방법은 얼굴의 특징을 빠르고 쉽게 얻을 수 있고, 다른 방법에 비해 계산량을 줄일 수 있다. 따라서 본 논문에서는 실시간의 전제조건을 만족하면서 강인성을 좀 더 보장할 수 있는 기법에 연구의 목표를 두었다.

본 논문의 구성은 2장에서는 장면전환 검출을 설명한다. 3장에서는 얼굴검출을 설명한다. 4장에서는 시스템 구현 및 시뮬레이션 한 결과를 보여주고, 5장에서는 결론을 맺는다.

## II. 장면전환 검출(Scene Change Detection)

동영상 데이터를 보다 효율적으로 관리하기 위해 인물을 포함하는 장면과 분할하여 구조화할 필요가 있다. 동영상 데이터의 계층적 구조는 크게 4개의 구조로 구성되는데 프레임(frame), 샷(shot), 씬(scene)

드), 비디오가 있다. 여기서, 동영상의 가장 하위 단위는 프레임(frame)으로 샷을 이루는 기본 단위이다. 그리고, 샷(shot)은 비디오 제작시 카메라의 기록과 멈춤 사이에서 얻은 일련의 연속적 프레임으로 비디오의 내용 해석을 위한 가장 기본적인 단위이다. 샷의 검출은 카메라 브레이크(camera break) 혹은 장면 전환검출(scene change detection)이라고 부르며 씬의 검출보다 비교적 쉬운편이다. 또한, 비디오의 바로 아래 단계는 씬(scene)으로 불리는 비디오의 이벤트(event) 단위의 시간축 분할에 해당되는 것으로 도서의 장(chapter)과 유사한 개념이다. 컴퓨터에 의해 비디오를 의미 있는 씬으로 나누는 것은 상당히 어려운 문제로 비디오의 핵심 단위인 샷을 클러스터링(clustering)하여 어느 정도 달성할 수 있다. 마지막으로, 동영상의 가장 상위구조는 비디오인데, 이는 제작자나 감독자에 의해 구별될 수 있는 단위로 도서에 책의 단위와 유사하다고 보면 된다.

장면전환은 연결 방식에 따라 급진적인 장면 전환과 점진적인 장면 전환으로 나눌 수 있다. 급진적인 장면 전환인 컷(cut)은 두 개의 장면들이 연속하여 연결되며, 점진적인 장면 전환은 두 개의 연속한 장면들이 한 프레임 이상 동안 서서히 전환하는 페이드(fade), 디졸브(dissolve) 등을 말한다. 가장 대표적인 급진적인 장면전환 추출에는 히스토그램 비교 방법과 화소단위 비교 방법이 있다<sup>[4]</sup>. 화소단위 비교 방법은 동일한 샷 내에서는 비교 화소 값의 변화가 적다는 성질을 이용하는 것으로, 식 (1)와 같이 인접하는 프레임의 대응하는 화소 값의 차이가 임계값  $t$ 를 초과하는 경우 '1'로 된다.

$$Dp_i(x, y) = \begin{cases} 1, & \text{if } |F_i(x, y) - F_{i+1}(x, y)| > t \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

여기서  $F_i(x, y)$ 는  $i$ 번째 프레임에서  $(x, y)$ 의 화소 값을 의미한다. 또한  $Dp$ 는 화소 값의 차이를 의미한다. 전체 화소에 대해 1로 된 화소의 비율이 특정 임계값  $T_p$ 를 넘으면, 컷으로 간주한다. 이 과정을 식(2)에 정의하였다.

$$\frac{\sum_{x, y=1}^{X, Y} Dp_i(x, y)}{X \times Y} \times 100 > T_p \quad (2)$$

여기서  $x, y$ 는 이미지의 최대 높이와 넓이를 표시한다. 이 방법은 구현이 간단한 반면, 카메라 연산과 재

체 움직임에 민감하여, 샷의 경계가 아님에도 불구하고 고 경계로 잘못 판정될 가능성이 크다.

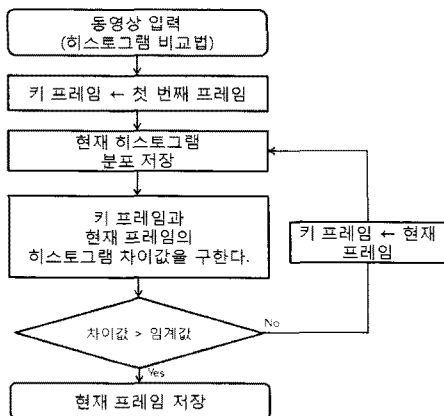
히스토그램 비교 방법은 동일한 샷으로 분류되는 프레임들의 색상분포는 유사하다는 성질을 이용하여 컷을 검출한다. 이 방법은 각 프레임의 컬러 히스토그램 차이를 계산한 후, 식(3)과 같이 그 차이 값이 임계값을 넘으면 컷으로 판단한다.

$$\sum_{j=1}^N |H_i(j) - H_{i+1}(j)| > T \quad (3)$$

여기서  $H_i(j)$ 는  $i$ 번째 프레임에서  $j$ 레벨의 히스토그램 값을 나타내고,  $N$ 은 히스토그램의 레벨 수를 나타낸다. 인접한 프레임들의 차이를 정해놓은 임계값  $T$ 와 비교함으로써 컷을 검출한다.

히스토그램 비교 방법은 구현이 쉽고 연산이 적어 효율적이며 움직임이 있는 장면에서도 전체적인 명암도는 크게 변하지 않으므로 물체의 움직임이나 카메라 이동에 덜 민감하다. 그러나 플래시 효과처럼 밝기가 순간적으로 변화하는 경우에는 대처하기 곤란하고, 다른 샷이지만 색상분포가 유사할 때에는 오검출이 발생하는 단점이 있다<sup>[5]</sup>.

따라서, 본 논문에서는 영화, 드라마와 같은 동영상은 프레임간의 유사성이 높으므로 움직임과 카메라 이동에 덜 민감한 히스토그램 비교 방법을 이용하여 장면전환 검출을 수행하여 얼굴검출을 하였다. [그림 1]에서와 같이 입력된 동영상을 필터링 한 후 각각의 컬러 히스토그램을 RGB 각각에 대하여 산출하고 키 프레임과 현재 프레임의 차이 값을 계산하여 임계값보다 클 경우 현재 프레임에 대해서는 영상의 재생 순서대로 저장을 하고, 임계값을 넘지 못했을 때는 장면



[그림 1] 장면전환 검출 순서도

전환이 이루어 졌는지 조사하고 저장하는 순서를 거치지 않는다. 여기서, 키 프레임은 첫 번째 프레임을 포함하여 추출된 각 샷들의 첫 번째 프레임이 된다. 또한 장면전환 검출 단계에서 얻어진 히스토그램의 차이 값에 따라 차이값 변화가 큰 부분이 전환점이 되는 프레임을 찾아 이 프레임을 키 프레임으로 설정하도록 한다.

### III. 얼굴 검출(Face Detection)

얼굴 검출 과정은 얼굴의 피부색을 이용하여 얼굴의 후보 영역을 검출, 얼굴 특징점의 색상 정보를 이용한 얼굴 특징점 검출로 이루어진다.

#### 3.1 얼굴 후보 영역 검출 (Face Localization Detection)

칼라 영상에서 피부색 정보는 얼굴 영역을 분리해 내기 위한 중요한 요소이다. 본 논문에서는 RGB 칼라 모델을 YCbCr 칼라 모델로 변환하여 색차 신호인  $C_b$ 와  $C_r$  성분만을 이용하여 피부색 영역을 검출한다. YCbCr 색공간은 식(4)와 같은 방정식으로 나타낼 수 있다<sup>[6]</sup>. 어둡거나 밝은 피부색의 검출에 보다 강건하게 된다.

$$\frac{(x - ec_x)^2}{a^2} + \frac{(y - ec_y)^2}{b^2} = 1, \quad (4)$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} C_b - c_x \\ C_r - c_y \end{bmatrix},$$

where  $c_x = 109.38$ ,  $c_y = 152.02$ ,  $\theta = 2.53$ (in radian)  $ec = 1.60$ ,  $ec = 2.41$ ,  $a = 25.39$ , and  $b = 14.03$  are computed from the skin cluster in the space

여기서, [그림 2]의 주어진 입력 영상 (a)에 대해  $c_b, c_r$  공간 안에서 피부색 영역분할 과정을 거친 결과가



(a) 입력 영상

(b) 색상 영역

[그림 2] 피부색 영역 추출

[그림 2]의 (b) 영상이다. [그림 2]의 (b) 영상을 보면 세밀한 잡음 형태가 존재하는 것을 볼 수 있다.

분할된 영역들은 얼굴뿐만 아니라 잡음 영역이 포함될 수 있다. 따라서 분할된 영역이 잡음 형태를 제거하기 위해서 형태학적 필터(morphological filter)의 식(6)에 정의된 제거(Opening)와 식(7)에 정의된 채움(Closing) 연산을 수행한 다음 침식(erosion) 연산식(8)과 팽창(dilation) 연산 식(9)을 수행한다.

$$r(f) = \delta(\epsilon(f)) = (f \ominus B) \oplus B \quad (6)$$

$$\phi(f) = \epsilon(\delta(f)) = (f \oplus B) \ominus B \quad (7)$$

$$c(f(x)) = \text{Min}\{f(x+k), k \in B\} = f \ominus B \quad (8)$$

$$\phi(f(x)) = \text{Max}\{f(x-k), k \in B\} = f \oplus B \quad (9)$$

위와 같은 형태학 연산을 적용한 영상의 단순화는 경계를 보존하면서 미세 정보들을 제거할 수 있어 영상 분할을 용이하게 할 수 있는데, 최근에 형태학 연산을 이용한 단순화 방법이 많이 사용되고 있다. [그림 3]은 얼굴 후보 영역으로 결정된 영상이다.



(그림 3) 얼굴 후보 영상

### 3.2 얼굴 특징점 검출(Face Feature Detection)

검출된 얼굴 영역 안에서 얼굴의 특징 가운데 눈과 입은 얼굴 인식을 위해 사용되는 대표적인 특징점이다. 이러한 눈과 입 검출을 위해서 형판(template), 에지 검출, 히스토그램 분석 등 다양한 방법이 제안되어 왔다<sup>(7)</sup>.

형판은 미리 정해 놓은 얼굴 특징에 관한 형판을 가지고 상관 관계를 조사하여 검출하는 방법이다. 형판을 이용하는 방법은 크기와 회전에 따른 형판을 많이 만들어 비교를 해야 하므로 찾는 시간이 많이 소요되며, 데이터를 저장해야 할 공간도 많이 필요하다. 에

지 검출 방법을 이용할 경우 에지가 조명에 많은 영향을 받고 잡음에 의해 손상되기 때문에 얼굴 특징을 검출하기 어렵다. 히스토그램 분석 방법은 눈이 얼굴 영역에서 어두운 분포를 가지고 있다는 것을 이용한 것으로 조명에 강인하지 못하다.

이러한 문제를 해결하고자 본 논문에서는 얼굴 특징점 검출 방법을 색차 성분 값을 이용하는 YCbCr 색상 공간에서 산출하는 식(10), (11)을 이용하였다.

$$Eye\ C = \frac{1}{3} \{ (C_b^e) + (\bar{C}_r)^2 + (C_b/C_r) \} \quad (10)$$

$$Mouth = C_r^e \cdot (C_r^e - \eta \cdot C_r/C_b)^2, \\ \eta = 0.95 \cdot \frac{\frac{1}{n} \sum_{(x,y) \in FG} C_r(x,y)^2}{\frac{1}{n} \sum_{(x,y) \in FG} C_r(x,y)/C_b(x,y)} \quad (11)$$

$C_b^e, (\bar{C}_r)^2, C_b/C_r$ 은 0에서 255사이의 값으로 정규화 되어 있으며,  $C_b, C_r$ 은 각각 YCbCr의  $C_b$ 와  $C_r$ 값이며  $(\bar{C}_r)^2$ 은 역변환된  $C_r$ 값이다.  $n$ 은 얼굴후보영역 마스크 FG의 화소개수이며,  $\eta$ 은  $C_r^e$ 의 평균값과  $C_r/C_b$  평균값의 비율이다.

눈 영역은 색차 성분 중 [그림 4]의 (a)와 같이  $C_b$  성분에서 높은 화소 값을 갖고,  $C_r$  성분과 휘도성분인 Y성분에서 낮은 화소 값을 갖는다. 그리고, 입 영역은 얼굴에서 붉은색 성분을 많이 포함하고 있으므로 [그림 4]의 (b)와 같이  $C_r$  성분에서 높은 화소 값을 갖고  $C_b$  성분과 휘도성분인 Y성분에서 낮은 화소 값을 갖는다.

이러한 얼굴 특징점이 부각되는 과정에서 실제 얼굴의 경우 눈과 입이 부각되는 특징을 가지게 되고 실제 주변 배경 화소값이 눈, 입보다 더 높은 값을 가질 수 있으므로 기하학적 특징을 이용하여 눈, 입 사이의 거리를 계산하여 그 거리가 비정상적일 경우 검출에서 제외시켰다.



(a) Cb영상(눈) (b) Cr영상(입)

(그림 4) 얼굴 특징점이 부각된 영상

효율적인 입 검출을 위해 높은 화소 값을 많이 갖도록 팽창연산(dilation operation)을 한다. 눈의 경우, 우선 입을 검출하므로 그 특징을 이용하여 눈을 부각 후 입의 위 부분만 검색한다. 한 개의 눈을 찾고 검출된 입의 위치와 비교하여 다른 눈을 찾는다. 다른 눈을 찾을 경우 입과 먼저 찾은 눈의 위치를 이용하여 검색 범위를 제한한다.

IV. 실험 결과







본 논문에서는 Microsoft Visual Studio C++ 6.0을 사용하였다.

실험 동영상으로는 다양한 얼굴 영역을 포함하고 있는 드라마 비디오를 이용하였다. 제안된 알고리즘의 성능 검증을 위해서 온전한 얼굴을 가진 장면전환 검출 영상과 이를 이용한 얼굴 검출에서 눈이 두 개 이상이고 입이 바르게 검출된 경우를 정상 얼굴로 판단한다.









여기서, 얼굴검출 영상에서 실제 얼굴이 정면영상 뿐만 아니라 옆면, 측면 얼굴까지 검출하는 것을 볼 수 있다. 이는 후보 영역 설정 시 고립점들을 모두 채워주고, 실제 후보 영역 모두를 사용하는 것이 아니라 이진화된 후보 영역만 사용하므로 배경 화소값으로부

터 얼굴의 특징점이 잘 부각 될 수 있다. 또한 얼굴의 기하학적 특징을 이용해 눈, 입 사이의 거리를 측정하여 그 거리가 상대적으로 비정상적인 경우 검출에서

(표 2) 얼굴 방향에 따른 미 검출된 얼굴

	원본	미검출된 얼굴
정면 영상		
정면에 가까운 영상		
측면에 가까운 영상		
측면 영상		

(표 1) 얼굴 방향에 따른 검출된 얼굴

	원본	검출된 얼굴
정면 영상		
정면에 가까운 영상		
측면에 가까운 영상		
측면 영상		

(표 3) 얼굴 방향에 대한 검출 성능 비교

얼굴방향	영상의 개수	검출률(%)	
		다른 알고리즘	제안 알고리즘
정면	48	89.40	91.32
정면에 가까운	30	90.74	89.21
반 옆면	21	74.67	79.06
측면	12	18.18	43.80



(그림 5) 등장인물 검출 프로그램

제외하였다.

검출된 입의 위치를 이용하여 눈 위치 영역을 제한하였더니 눈 검출 가능성을 기존 알고리즘보다 약 30% 높일 수 있었다. 얼굴 방향에 대한 검출 성능 비교는 [표 3]과 같다.

## V. 결 론

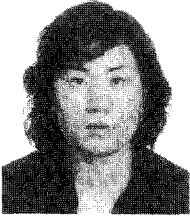
본 논문에서는 다수의 인물이 존재하는 드라마, 영화와 같은 비디오 동영상을 입력으로 사용하여 프레임 간의 히스토그램 비교를 통해 장면을 검출하였다. 검출된 장면에서 얼굴을 검출하기 위해 칼라 영상에서 RGB 영상을 YCbCr 좌표계로 변환한 후, 피부색 정보를 이용하여 얼굴 후보 영역을 추출하였다. 얼굴 후보 영역 내에서 얼굴의 특징점인 눈과 입을 검출하기 위해 얼굴의 색차 성분의 특징을 이용하였다.

제안된 알고리즘은 입의 위치를 이용하여 눈 위치 영역을 검출하므로써 [표 1]에서와 같이 정면 얼굴과 정면에 가까운 얼굴에서는 검출률이 약 87%로 비슷한 성능을 보이나 측면에 대해서는 43%로 높은 검출률이 보였음을 알 수 있다. 측면영상은 눈과 입이 1개씩 나오기 때문에 얼굴 후보영역을 설정할 때 부수적으로 발생하는 고립점을 이용하여 얼굴영역을 최대로 설정함으로써 측면에서 얼굴 특징점이 포함되도록 하였다.

## 참고문헌

- [1] Young Ho Kwon, da Victoria Lobo, "Face Detection using templates," Proceedings of the 12th IARR International conference on, Vol. 1, pp. 764-767, Oct. 1994.
- [2] H. A. Rowley, S. Baluja and T. Kanade, "Neural network based face detection", IEEE Trans Pattern Analysis and Machine Intelligence, vol 20, pp. 23-38, 1998.
- [3] H. Schneiderman and T. Kanade, "A Statistical Method for 3D Object Detection Applied to Faces and Cars," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 746-751, June 2000.
- [4] Ronghua Liang, Chun Chen, Jiajun Bu, "Real-time facial features tracker with motion estimation and feedback", IEEE Systems, Man and Cybernetics Conf., vol. 4, pp.3744-3749, Oct. 2003.
- [5] S.Z. Li, L. Zhu, Z.Q. Zhang, and H.J. Zhang, "Statistical Learning of Multi-View Face Detection," In Proc. 7th European Conference on Computer Vistion, Copenhagen, Denmark. May 2002.
- [6] 백승호, "비디오 등장인물 검색을 위한 얼굴검출", 석사학위논문, 광운대학교, 2006년 2월.
- [7] 김영구, 노진우, 고한석, "복잡한 배경의 칼라영상에서 Face and Facial Feature 검출", 대한전자공학회 하계종합학술대회 논문집, 25(1), pp.69-72, 2002년 6월
- [8] 김기남, 김형준, 김희율, "방송 비디오 등장인물 자동 분석 시스템", 방송공학회논문지 9(2), pp.164-173, 2004년 6월
- [9] 나중원, 강대욱, 배중성, "색상정보와 AdaBoost 알고리즘을 이용한 얼굴검출", 한국해양정보통신학회논문지 12(5), pp.843-848, 2008년 4월.

〈著者紹介〉

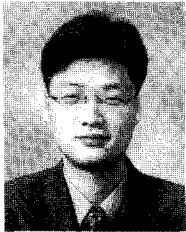


조 미 남 (Mi-Nam Cho)

2002년 2월: 동신대학교 정보통신공학과 공학석사

2009년 2월: 동신대학교 정보통신공학과 공학박사

〈관심분야〉 RFID시스템, 데이터베이스, 영상처리



지 유 강 (Yoo-Kang Ji)

2006년 2월: 동신대학교 정보통신공학과 공학박사

2006년 3월~2009년 8월: 동신대학교 정보통신공학과 전임강사

〈관심분야〉 영상통신, 네트워크보안, 임베디드시스템, RFID/USN