제스쳐 인식을 이용한 DID 인터페이스 구현

이상헌*, 김대진**, 최홍섭***

요 약

본 논문에서는 DID 시스템에서 사용할 수 있는 제스처 인식을 이용한 비접촉식 인터페이스를 구현하였다. 비접촉식 인터페이스는 별도의 부착물 없이 키넥트 카메라만을 사용함으로, 사용자의 편의와 공간적인 활용도를 높일 수 있다. 손 동작인식에는 사용자의 손 움직임의 기울기와 속력을 인식하는 방향성 기반의 인식 기법을 채용하였고 손 모양인식을 위해서 YCbCr 칼라모델을 이용한 손 영역 추출과 손 넓이의 원을 이용한 영상처리 기술로 손가락의 수를 인식하였다. 이러한 손 동작인식과 손 모양인식을 이용하여 다음 페이지, 이전 페이지, 화면 위로, 화면 아래로, 커서 움직임, 클릭 등의 이벤트를 발생시켜 DID 시스템 제어 명령으로 사용하였으며, 구현한 시스템을 갖고 동작 실험한 결과 93%의 명령인식률을 보여 실용화의 가능성을 확인할 수 있었다.

Implementation of DID interface using gesture recognition

Sang-Hun Lee*, Dae-jin Kim**, Hong-Sub Choi***

Abstract

In this paper, we implemented a touchless interface for DID(Digital Information Display) system using gesture recognition technique which includes both hand motion and hand shape recognition. Especially this touchless interface without extra attachments gives user both easier usage and spatial convenience. For hand motion recognition, two hand-motion's parameters such as a slope and a velocity were measured as a direction-based recognition way. And extraction of hand area image utilizing YCbCr color model and several image processing methods were adopted to recognize a hand shape recognition. These recognition methods are combined to generate various commands, such as, next-page, previous-page, screen-up, screen-down and mouse -click in oder to control DID system. Finally, experimental results showed the performance of 93% command recognition rate which is enough to confirm the possible application to commercial products.

Key words: gesture recognition, DID, Kinect camera, hand motion recognition, hand shape recognition

1. 서 론

디지털 정보 디스플레이(DID, Digital Information Display) 시스템은 기존의 수작업에

※ 제일저자(First Author): 이상헌 접수일:2012년 08월 07일, 수정일:2012년 09월 05일

** (주)다이렉트미디어

완료일:2012년 09월 10일

- *** 대진대학교 전자공학과 교수(교신저자)
- 이 논문은 2012학년도 대진대학교 학술연구비 지 원에 의한 것임

기반을 둔 게시판을 대치할 새로운 디스플레이 장치로서 주목받고 있다. 이러한 DID는 기존의 게시판과 달리 다양한 형태의 디스플레이가 가능하며 일방적인 정보전달이 아닌 문서의 편집, 저장, 관리, 검색 등의 컴퓨터 기반의 다양한 제어가 가능하기 때문에 지하철, 버스 정류장, 공항 등의 공공장소에서의 정보제공 및 백화점과 쇼핑몰 등 매장에서의 광고 장치로 많이 사용되는 추세이다. 하지만 현재 DID는 일방적인 정보제공에 그치는 경우가 대부분이며, 이는 정보의 단방향 흐름에만 의존함으로 그 활용도가 매우 떨어지고 있다. 따라서 DID 시스템에 직관적이며 효율적인 인터페이스 기술개발의 필요성

^{*} 대진대학교 전자공학과 석사과정 kl0525@hanmail.net

이 커지고 있어서 본 논문에서는 별도의 부착물 및 장비 없이 동작인식을 이용한 비접촉식인터페이스방식을 제안하여 사용자의 편의성을 높이고자 하였다.

기존의 DID 시스템의 경우 터치스크린을 이용한 인터페이스가 대부분으로 사용자에게 익숙하고 인식률도 높아 많은 분야에서 활용되고 있다. 하지만 적외선 방식과 초음파 방식을 사용하는 DID시스템의 대형 스크린은 높은 비용과센서 오염에 취약한 단점을 가지고 있으며, 또다른 방식인 감압식과 정전식은 소형 스크린에 적용할 수 있기 때문에 DID시스템에서는 적절하지 않다.

본 논문에서는 DID 시스템 인터페이스를 위하여 손 움직임 속도와 방향 정보를 이용한 손동작인식 알고리즘과 YCbCr 칼라모델과 영상의 팽창과 침식을 이용한 손모양 인식 알고리즘을 구현하였다[1-4].

본 논문의 구성은 제2장에서는 손 동작인식과 손모양 인식 방법에 관해서 설명하며, 제3장에서는 제안된 방식의 실험결과와 분석을 그리고 마지막으로 제4장에서는 결론 및 향후 연구에 관해 기술한다.

2. 연구내용

본 논문에서는 DID 시스템의 인터페이스를 구현하기 위해 손 동작인식과 손 모양인식을 이용하였으며 시스템의 동작 블록도는 그림1과 같다.



(그림 1) DID 인터페이스 시스템 동작 블록도

2.1 손동작 인식

입력 장치로 이용되는 키넥트 카메라는 영상

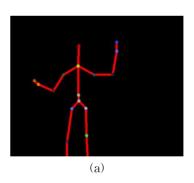
의 깊이 정보와 컬러 정보와 더불어 사용자 관절의 3차원 좌표까지 제공해준다. 키넥트 카메라를 통해 입력된 위와 같은 정보를 이용하여 손동작인식을 하였으며, 과정은 아래 (그림2)와 같이 5단계로 구성된다.



(그림 2) 손동작 인식 5단계

2.1.1 손 추적 및 기울기 속도 계산

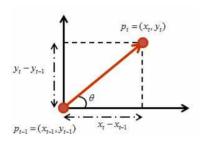
(그림 3)은 키넥트에 의해 제공되는 인체의 관절 정보를 보여주고 있다. 이를 이용하여 동 작인식을 하면 깊이 정보와 왼손과 오른손을 구 분할 수 있는 장점이 있다. 본 논문에서는 주로 손의 관절 정보를 사용하며 어깨와 몸통 정보를 통해 상대적인 위치를 알 수 있도록 했다.





(그림 3) (a) 키넥트를 이용한 주요 관절 추적 (b) 실제 사용자의 영상

(그림 4)는 손 관절의 움직임을 좌표로 보여주는 그림으로 $p_t=(x_t,y_t)$ 는 시간 t에서의 손 관절의 좌표이며, $p_{t-1}=(x_{t-1},y_{t-1})$ 는 시간 t-1에서의 좌표를 나타낸다.



(그림 4) 손동작의 기울기와 속도 계산

손 동작 인식을 위해 사용자 손 동작의 기울 기 및 속도를 아래 식(1)을 이용하여 계산한다 [1][4].

$$d = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}$$

$$\theta = \tan^{-1} \left(\frac{y_t - y_{t-1}}{x_t - x_{t-1}}\right)$$

$$v_t = d \times f_{ps}$$
(1)

d는 한 프레임 동안 움직인 거리이며, θ 는 움직임의 기울기를 말한다. 여기서 거리 d에 프레임율 (fps: frame per sec)를 곱하면 속력을 구할 수 있다.

2.1.2 동작인식 및 이벤트 발생

본 논문에서 인식되는 동작은 (그림5)와 같이 4가지 종류이며, 매우 단순하면서도 동작의 방향성이 확실하게 구분된다.



(그림 5) 동작 인식에 사용된 동작의 4가지 방향성

동작 구간을 프레임 단위로 1부터 n까지라하고, 동작 구간내의 한 프레임에서의 손동작의기울기 θ 는 식(2)에 의해 4가지 방향을 기본 벡터 a, b의 형태로 변환하며, 이 방향성을 그 프레임의 동작 속력과 곱을 한 값들의 평균 A를 식(3)을 이용하여 구한다.

$$.\alpha = \begin{cases} b & (45^{\circ} < \theta \le 135^{\circ}) \\ -a & (135^{\circ} < \theta \le 225^{\circ}) \\ -b & (225^{\circ} < \theta \le 315^{\circ}) \\ a & (315^{\circ} < \theta \le 45^{\circ}) \end{cases}$$
(2)

$$A = \frac{\sum_{i=1}^{n} (v_i \alpha_i)}{n} \tag{3}$$

이렇게 구해진 A는 식(1)과 같은 방법으로 동작구간의 기울기와 크기를 구할 수 있으며, 이 값들을 문턱치 조건을 통해 <표 1>과 같은 이벤트를 발생시킨다.

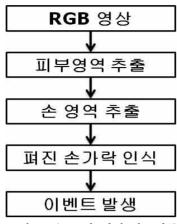
<표 1> 동작인식과 이벤트 발생

발생 이벤트	
다음 페이지	
이전 페이지	
아래로	
위로	

2.2 손 모양 인식

손 모양 인식 알고리즘의 과정은 (그림6)과

같으며 이를 통해 동작구간 추출, 마우스 커서 움직임, 클릭 등의 이벤트를 발생시킨다.



(그림 6) 손 모양 인식 알고리즘

2.2.1 손 영역 추출

손 영역을 추출하기 위해서는 우선 입력영상에서 피부영역을 추출해야 한다. 피부영역은 RGB 컬러 영상을 YCbCr 컬러 영상으로 변환하여 피부색을 구하게 된다. RGB 모델에서 YCbCr 모델로의 변환은 식(4)와 같은 연산을이용하며, 이렇게 변환된 영상에서 Cb와 Cr값이 식(5)의 조건을 만족하는 영역을 피부색으로구분하여 추출하게 된다[3].

$$Y = 0.2989R + 0.5866G + 0.1145B$$

$$Cb = -0.16874R - 0.33126G + 0.5000B$$

$$Cr = 0.5000R - 0.41869G - 0.0816B$$
(4)

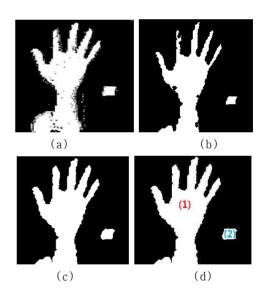
$$77 < C_b < 127$$
 (5)
$$140 < C_r < 173$$

(그림 7)의 (a)는 RGB 영상이며 (b)는 위와 같은 방법으로 피부영역을 추출해낸 결과의 영 상을 보여준다.



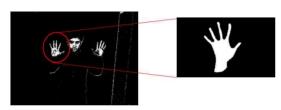
(그림 7) (a) RGB 영상 (b) 피부영역 검출 결과 영상

이렇게 구해진 피부색 영상에는 일반적으로 잡음이 산재해 있기 때문에 팽창 및 침식 영상처리 기법과 외각선 추출을 통한 라벨링을 이용하여 잡음을 제거하는 과정을 거친다. (그림 8)(a), (b), (c)는 원 영상에서 침식과 팽창을 통해 잡음이 제거되는 과정을 보여준다. 또한 이렇게 잡음이 제거된 (c)영상에서 외곽선 추출을통해 (d)와 같이 라벨링 작업을 한 후 일정 크기 이하는 잡음으로 처리하여 제거한다. 그림 9은 잡움 제거를 하기 전 영상과 제거한 후 그로부터 손 영역만을 추출한 결과 영상을 보여준다[3].





(그림 8) (a) RGB 영상 (b) 침식 연산 후 영상 (c) 팽창 연산 후 영상 (d) 라벨링 영상 (e) 최종 잡음 제거 영상



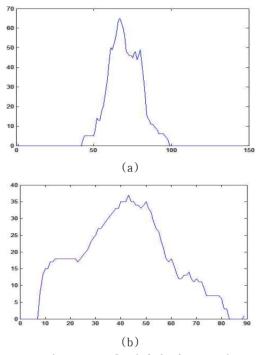
(그림 9) 잡음제거와 영역 추출

다음으로 추출된 손 영역의 영상에서 손목 아래 부분을 제거하기 위해 손 영역의 외접 사 각형의 중심점과 손 면적의 무게 중심을 구한 다.

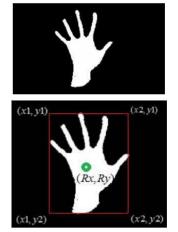
외접 사각형의 중심점을 구하기 위해서 영상의 수평히스토그램과 수직 히스토그램을 이용한다. (그림 10)와 같이 히스토그램의 시작 부분 (x1,y1)과 마지막 부분(x2,y2) 사이의 사각형 영역 내에서만 손의 이미지가 존재하며, 영역의중심 좌표는 식(6)을 통해 구할 수 있으며, (그림 11)은 손 영역의 외접 사각형과 그 중심점좌표를 보여준다.

$$Rx = \frac{x1+x2}{2}$$

$$Ry = \frac{y1+y2}{2}$$
(6)



(그림 10) (a) x축 방향의 히스토그램 (b) y축 방향의 히스토그램



(그림 11) 손 영역의 외접 사각형과 중심점

손 영역의 무게 중심은 식(7)과 (8)을 이용한다. 여기서 C(i,j)는 좌표 (i,j)의 영상 값이되며, 흰 부분은 1, 검은 부분은 0의 값을 갖는다. B는 손 영역의 면적을 뜻하고, (Dx,Dy)는 손 영역의 무게 중심점이 된다.

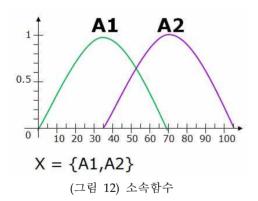
$$B = \sum_{i=x_1}^{x_2} \sum_{j=y_1}^{y_2} C(i,j) \tag{7}$$

$$Dx = \frac{\sum_{i=x1}^{x2} \sum_{j=y1}^{y2} jC(i,j)}{B}$$

$$Dy = \frac{\sum_{i=x1}^{x2} \sum_{j=y1}^{y2} iC(i,j)}{B}$$
(8)

2.2.2 min-max 중심방법 퍼지 제어 알고리즘

퍼지 추론이란 불분명하고 모호한 상태를 참혹은 거짓의 이진 논리가 아닌 다치성으로 표현하는 논리 개념을 말한다. 퍼지 논리는 근사치나 주관적 값을 사용하는 규칙들을 생성함으로써 부정확함을 표현할 수 있는 규칙 기반 기술로 본 논문에는 Mamdani의 min-max 중심법을사용하였다. 먼저 퍼지 제어 시스템의 입력 변수 X와 Y에 대한 퍼지(상태)집합으로 각각{A1, A2}, {B1, B2}가 있으며, 상태 Ai와 Bi에대한 소속함수는 (그림12)와 같은 모양을 가졌으며, 논의의 편의상 Ai와 Bi 모두 같은 모양이라고 가정한다.[3][8]



다음으로 퍼지 제어 시스템의 출력변수 Z에 대한 퍼지집합은 {C1, C2} 이라 하고, 입력상태와 출력상태의 연결 관계를 정의한 퍼지규칙은 아래 <표 2>와 같다.

<표 2> 퍼지규칙

	A1	A2
B1	C1	C2
B2	C1	C2

즉, 입력상태 A1과 B1에 의해 출력상태 C1이 결정되고, A2과 B1에 의해 출력상태 C2가 결정되는 관계에 있다. 구체적인 $C_n (n=1,2)$ 값은 아래 식(9)에 의해 구하게 된다

$$C_n^{n e w} = Min(\mu A_i(x), \mu B_j(y))$$

$$C_n = Max(C_n^{n e w}, C_n)$$

 $_{end}^{end}$

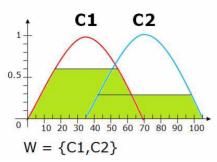
 $\mu A_i(x)$ = 입력값 x에 대한 Ai의 소속함수 출력값

 $\mu B_i(y)$ = 입력값 y에 대한 Bi의 소속함수 출력값

 n_A = X의 퍼지집합의 상태 개수

 n_B = Y의 퍼지집합의 상태 개수

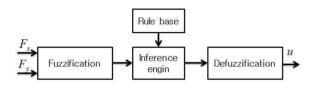
입력변수 x와 y의 값이 정해지면 그에 대한상태 Ai와 Bi의 Z합에 대해 소속함수의 출력 값을 계산한 후, 그중 작은 값을 C_n^{new} 으로 하고, 이를 기존의 C_n 값과 비교하여 큰 값을 최종 C_n 으로 정한다. 이와 같이 C1과 C2가 결정되면, 퍼지제어의 최종 출력 Z값을 구하기 위하여 역으로 비퍼지화를 하게 되는데 논문에서는 일반적으로 사용되는 무게중심법을 이용하였다. 무게중심법은 위에서 구한 C1과 C2값을 가지고 역으로 C_n 의 소속함수 면적의 무게중심을 계산하여 이를 퍼지제어 알고리즘의 최종 출력 Z로 결정하는 방식이다. (그림 I3)은 C1의 값이 I0.6 I1의 값이 I2이다.



(그림 13) 비퍼지를 위한 무게중심법

2.2.3 퍼지 제어를 이용한 손목 영상 제거

아래 (그림 14)는 손목제거를 위한 퍼지 제어 시스템의 구성도이다. 퍼지 제어 시스템의 입력 신호 Fx와 Fy는 앞의 과정에서 추출된 (Rx,Ry), (D_x,D_y) 값으로부터 식(10)을 사용하여 계산한다.[8]

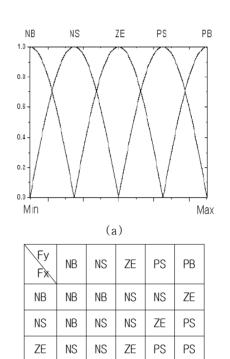


(그림 14) 퍼지 제어 시스템

$$Fx = Dx - Rx$$

$$Fy = Dy - Ry$$
(10)

입력 Fx,Fy에 대한 퍼지 집합은 PB(Positive Big), PS(Positive Small), ZE(Zero), NS(Negative Small), NB(Negative Big)와 같이 정의하였고, 그림15는 그에 대한 소속함수와 퍼지규칙을 나타낸다. 이때 소속함 수의 입력값의 범위는 Fx는 Min값은 -4, Max값은 16, Fy는 Min값은 -18, Max값은 6, 출력 u의 Min값은 -40, Max값은 40으로 하였다[2]. 무게중심법을 사용한 비퍼지화로 구한 최종 퍼 지 제어 출력값 u를 사용하여 해당 영상의 손목 여부를 판단하여 손 모양 이외의 부분을 영상에 서 제거하였다.



(b) (그림 15) (a) 소속 함수 (b) 퍼지 규칙

PS

PS

PS

PB

2.2.4 손 모양 인식 방법

PS

PB

NS

ZΕ

ZΕ

PS

카메라와 사용자의 거리가 일정하지 않기 때문에 손의 면적을 이용하여 원을 생성한다. 즉 손의 무게중심점으로부터 손의 면적에 해당하는 원을 생성하는데, 이때 원의 반지름 r은 다음과 같다.[3]

$$r = k\sqrt{\frac{B}{\pi}} \tag{11}$$

여기서 k는 비례상수 값을 의미하며, B는 손 영상의 면적이다. 이렇게 구해진 원을 기준으로 0°부터 360°까지 시계방향으로 돌면서 영상이 검은색으로부터 흰색으로 3개 이상 연속되는 경우의 수를 센 후, 여기서 하나를 빼주면펴진 손가락의 개수를 구할 수 있게 된다. 이렇게 인식된 손가락의 수에 따라 표 3와 같이 이벤트를 발생시키게 되며 그림 13은 여러 가지손 모양 인식의 예를 이미지로 보여 주고 있다.

<표 3> 손모양 인식과 이벤트 발생

펴진 손가락 수	발생 이벤트
0	마우스 누름
1~2	마우스 움직임
4~5	마우스 올림







(그림 16) 손 모양 인식의 예

3. 실험 및 평가

3.1 실험 환경

본 논문에서 컴퓨터 제어 인터페이스 구현을 위하여 OS로는 윈도우 7(Windows 7)을, 프로 그램 개발에는 openNI, OpenCV, OpenGL 등의 open Library를 사용하였다. 또한 Kinect SDK 의 유져 트래킹, 뼈 구조 생성 소스 등을 참고 하였으며, 컴파일러는 Microsoft Visual Studio 2010을 사용하였다.

<표 4> 시스템 개발 환경

	상세	l내역	
0S	- 윈도우 7		
open library	- OpenCV		
	- OpenNI		
	- OpenGL		
SDK	- Kinect SDK		
컴파일러	- Microsoft	Visual	Studio
ㅁ피근디	2010		

3.2 실험 결과 및 구현 화면

위와 같은 환경에서 연구실 인원 8명이 각각 100번씩 제스쳐 인식 제어 명령을 실험하였으며, 그 결과는 <표 5>와 같다. 여기서 기존 방식이란 가속도 센서를 이용한 동작인식률을 말한다[4].

<표 5> 실험 결과

	명령의 종류	인식률	기존방식
손 동작 인식	다음 페이지	95.2%	
	이전 페이지	94.6%	
	아래로	92.0%	92.05%
	위로	93.2%	92.03%
손 모양	커서 움직임	96.0%	
인식	클릭	91.2%	

기존 방법의 인식율이 92%에 비해 제안한 제스쳐 인식 방법은 평균 인식율이 93.7% 정도로 상용화 가능성이 충분함을 보여주고 있다. 그리고 실제 이미지 출력과 동작을 시연하기 위해서이미지 뷰어 프로그램(pictomio)에 적용하여 실제 출력된 화면과 동작되는 상황을 (그림 17)과 (그림 18)에서 볼 수 있다.





(그림 17) 화면 제어





(그림 18) 실제 동작

4. 결론

본 논문에서는 DID 시스템에서 사용할 수 있는 비접촉식 인터페이스를 구현하였으며, 이를 위해 손 동작 인식과 손 모양 인식기술을 사용하였다. 시스템 구현을 위한 입력 장치로 키넥트 카메라를 사용하였으며, RGB영상을 입력 받기 위해 openNI와 Kinect SDK를 사용하였다.

손동작 인식에서 인식하고자 하는 동작들은 직관적이며 방향성이 강한 동작으로, 동작의 기울기와 속력을 이용한 인식 방법을 사용한다. 또한, 손 모양 인식을 위해서 YCbCr 칼라 모델을 이용한 손 영역 추출과 퍼지 제어를 이용하여 손목 영상을 제거를 하였고, 손 넓이의 원을이용한 영상 처리 기술로 손가락의 수를 인식할수 있었다.

실험 결과 명령 인식률은 93.7% 정도로 나왔으며, 사용자가 별도의 부착물이나 장치 없이도 직관적이고 편리한 인터페이스로 활용할 수 있음을 확인할 수 있었다. 향후 연구방향은 시스템의 인식률을 높이기 위해 키넥트의 신체 주요관절 데이터의 잡음처리와 보다 많은 인식 동작을 추가함으로써 다양한 분야에 적용이 가능하게 하는 것이다.

참 고 문 헌

- [1] 김종호, 윤요섭, 김태영, 임철수 "은닉 마르코프 모델 기반 동작 인식 방법" 멀티미디어학회 논문지 제12 권 제4호 2009. 4
- [2] 황문구, 김형래, 강석범, 양태규, "퍼지 추론을 이용한 비전 기반 실시간 손모양 인식" 한국정보기술학회논 문집 제6권 제2호 2008. 2
- [3] Calibration of a Depth and Color Camera Pair", Proc eedings of the 14th International Conference on Co mputer Analysis of Images and Patterns (CAIP) 20 11, LNCS 6855, pp. 437–445.
- [4] 김상기, 박건혁, "3차원 가속도 데이타를 이용한 HM M 기반의 동작인식" 정보과학회논문지 컴퓨팅의 실제 및 레터 제 15 권 제 3 호 2009.
- [5] 임성빈, 최우경, "관성 센서를 이용한 동작 인식 시스템의 개발" Proceedings of KFIS Spring Conference M 16 권 제 1 호 2006.
- [6] 이상헌, 김대진, 최홍섭, "키넥트의 좌표변환을 이용한 동작인식과 그 응용에 관한 연구" 2011년 음향학회 추계 학술대회 논문집, 제30권, 2011
- [7] 이상헌, 김대진, 최홍섭, "키넥트를 이용한 동작인식 과 컴퓨터 제어 인터페이스 설계 "2011년 음향학회 춘계 학술대회 논문집, 제 30권, 2011.5
- [8] J. Jantzen, "Foundations of Fuzzy Control", John Wil ey & Sons, 2007.



이 상 헌 2011년 : 대진대학교 전자공학과 졸업(공학사)

2011년 [~] 현재 : 대진대학교 전자 공학과 석사과정

관심분야: 영상 처리, 음성신호처리, 패턴인식

김 대 진

1998년 : 대진대학교 전자공학과

졸업(공학사)

2000년 : 동국대학교 전자공학과

졸업(공학석사)

2008년 : 대진대학교 전자공학과

졸업 (공학박사)

2000년~2003년: 한빛소프트 주임연구원

2003년~2007년: 모토로라 코리아 전임연구원

2007년~2008년: 아이비인터넷 부장 2008년 ~2010년: 미디어웹 책임연구원 2010년 ~현 재: 다이렉트미디어 연구소장

관심분야: 저작권 보호, 멀티미디어 시스템, 디지털

콘텐츠, 멀티미디어 검색, IPTV 등

최홍섭

1985년 : 서울대학교 전자공학과

졸업(공학사)

1987년 : 서울대학교 전자공학과

졸업(공학석사)

1994년 : 서울대학교 전자공학과

졸업(공학박사)

1995년 ~ 현재: 대진대학교 전자공

학과 교수

관심분야: 통신 및 신호처리, 음성처리, 멀티미디어

시스템, IPTV 등