

## 형태소 분석 기반 전자책 수화 번역 프로그램

한솔이 · 김세아 · 황경호\*

### E-book to sign-language translation program based on morpheme analysis

Sol-Ee Han · Se-A Kim · Gyung-Ho Hwang\*

Department of Computer Engineering, Hanbat National University, Daejeon 34158, Korea

#### 요 약

인터넷의 발전과 스마트 디바이스의 확산으로 e-book에 대한 수요가 늘고 있다. 그러나 청각 손실로 정확한 언어를 배우기 어려운 청각장애인은 텍스트로만 이루어진 e-book 서비스를 사용하기 어렵다. 본 논문에서는 e-book의 문장을 읽어 수화 동영상으로 제공하는 안드로이드 기반 애플리케이션 프로그램을 설계하고 구현하였다. e-book의 한국어 문장을 수화 언어로 번역하기 위해 형태소 분석에 기반한 알고리즘을 사용하였다. 제안한 알고리즘은 3단계로 구성된다. 1단계는 수화 표현을 위한 문장 요소 제거, 2단계는 수화 표현의 변환 및 시제 표현, 3단계는 수화 높임말 용어 변경 및 위치 이동이다. 또한 수화 번역 품질에 대한 평가 방안을 제시하고 100개의 기준 문장에 대해 제안한 알고리즘을 통한 번역 결과의 우수성을 확인하였다.

#### ABSTRACT

As the number of smart devices increases, e-book contents and services are proliferating. However, the text based e-book is difficult for a hearing-impaired person to understand. In this paper, we developed an android based application in which we can choose an e-book text file and each sentence is translated to sign-language elements which are shown in videos that are retrieved from the sign-language contents server. We used the Korean sentence to sign-language translation algorithm based on the morpheme analysis. The proposed translation algorithm consists of 3 stages. Firstly, some elements in a sentence are removed for typical sign-language usages. Secondly, the tense of the sentence and the expression alteration are applied. Finally, the honorific forms are considered and word positions in the sentence are revised. We also proposed a new method to evaluate the performance of the translation algorithm and demonstrated the superiority of the algorithm through the translation results of 100 reference sentences.

**키워드** : 수화, 번역, 스마트폰 애플리케이션, 전자책, 형태소 분석

**Key word** : sign-language, translation, smart phone application, e-book, morpheme analysis

Received 17 November 2016, Revised 18 November 2016, Accepted 28 November 2016

\* Corresponding Author Gyung-Ho Hwang(E-mail:gabriel@hanbat.ac.kr, Tel:+82-42-821-1751)

Department of Computer Engineering, Hanbat National University, Daejeon 34158, Korea

Open Access <http://doi.org/10.6109/jkice.2017.21.2.461>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.  
Copyright © The Korea Institute of Information and Communication Engineering.

## I. 서론

인터넷의 발전과 스마트 디바이스의 확산으로 인해 e-book 콘텐츠와 서비스가 많이 제공되고 있다. 그러나 문맹률이 30%에 이르는 청각장애인은 스마트 디바이스의 콘텐츠 정보의 접근성이 매우 떨어진다[1]. 청각 손실로 인해 정확한 언어를 배우기 힘들며 한국어와 수화 언어의 문법차이로 텍스트로만 이루어진 e-book의 일반적인 독서가 어렵다.

청각장애인에게 도움을 주기위해 현재 여러 서비스가 제공되고 있다. 국립장애인도서관에서 운영하는 장애인 정보 누리터는 장애인 이용자들이 도서관 자료에 쉽게 접근할 수 있도록 지원한다. 청각장애인에게는 진동이자, 화상전화기, 화상캠 등의 보조공학기기를 배치하고 있으며 전자도서를 이용하여 디지털 도서 형식으로 컴퓨터나 전용단말기를 통해 열람할 수 있도록 하고 있다. 또한 국립장애인도서관 영상자료관 스마트폰 애플리케이션을 개발하고 책을 수화로 구현한 영상을 제작하여 제공하고 있다.

그러나 국립장애인도서관에 직접 가서 편의 시설을 활용하기 어려운 청각장애인이 많고, 애플리케이션 서비스는 책의 내용을 사람이 수화로 직접 구현하는 것을 녹화하여 제공하는 것이기 때문에 제작하는데 비용이 많이 드는 문제점이 있다.

본 논문에서는 이와 같은 문제점들을 고려하여 청각 장애인들이 스마트 디바이스에서 쉽게 사용할 수 있도록 구성을 단순하게 하고 전자책의 내용을 수화동영상으로 보여줄 수 있어 글에 대한 접근성을 높이는 스마트폰 애플리케이션을 구현한다.

본 논문은 다음과 같이 구성되어 있다. II장에서는 한국어 문장을 수화문장으로 번역하는 기존 연구에 대해서 살펴본다. III장에서는 한국어 문장을 수화 언어로 번역하는 알고리즘을 제안한다. IV장에서는 수화로 번역된 언어에 대한 번역 성능을 평가하기 위한 자체 평가표를 제시하고 이를 바탕으로 제안한 알고리즘의 성능을 살펴본다. V장에서는 III장에서 제안한 수화 번역을 사용하여 전자책의 한국어 문장을 수화 동영상으로 보여주는 안드로이드 기반 애플리케이션으로 구현한 결과를 보여주고, VI장에서는 본 논문의 결론을 맺는다.

## II. 관련연구

청각장애인이 사용하는 수화 언어는 한국어와는 다른 언어체계를 가지고 있다. 이와 같은 언어의 차이점으로 인해 소통의 어려움이 있다. 한국어 문장과 수화 언어의 차이를 고려하여 한국어 문장을 수화 언어로 번역하는 연구가 많이 진행되고 있다.

[2]에서는 수화에서의 표현 특징을 살펴보고, 이를 바탕으로 한국어 복문 구조를 분석하여 청각 장애인들이 이해하기 쉬운 형태로 변환하여 자연스러운 수화를 생성하는 수화 액션스크립트 생성 시스템을 제안하였다.

[3]에서는 수화에서 나타나는 언어 현상을 재현, 생략, 변형, 이동의 네 가지 관점에서 정리했다. 그리고 정리한 관점을 결합범주문법에 반영하여 수화 문장을 처리하는 방법을 제안하였다.

[4]에서는 한국어-한국 수화 번역을 위한 규칙 기반의 기계번역 방식을 제안한다. 그러나 문형에 따라 인위적으로 작성하여 실제 사용하는 각 문형들이 고려되지 않은 단점이 있다.

[5]에서는 기존에 구축된 한국어-한국 수화 시스템에서 시제 표현의 잘못된 부분을 지적하여 한국어 문장을 적절한 수화 문장으로 변환하기 위해 필요한 시제 표현 방식을 제안하였다.

기존의 수화 번역 알고리즘을 종합적으로 고려하여 본 논문에서는 형태소 분석에 기반을 둔 새로운 방법을 제안한다.

## III. 수화 번역 알고리즘

본 논문에서 제시하는 수화 번역 알고리즘은 형태소 분석 과정을 거친 후 수화 언어로 번역하는 방법을 사용한다. 그림 1은 형태소 분석과 수화 번역에 대한 구성도이다. 형태소 분석기는 Komoran 2.0을 사용하였다 [6]. 형태소 분석기 단계 순서는 Komoran 라이브러리를 참조하였다. 형태소 분석기는 문장이 입력되면 자소변환을 하게 된다. 그 후에 TRIE 알고리즘 탐색을 통해 입력된 문장의 여러 분석을 통해 형태소들을 찾게 된다. 찾은 형태소들은 Lattice 형태의 자료구조에 저장되고 생성한다. 생성된 Lattice는 품사 태깅을 통해 문장에 맞

는 형태소를 찾아 결과를 낸다.

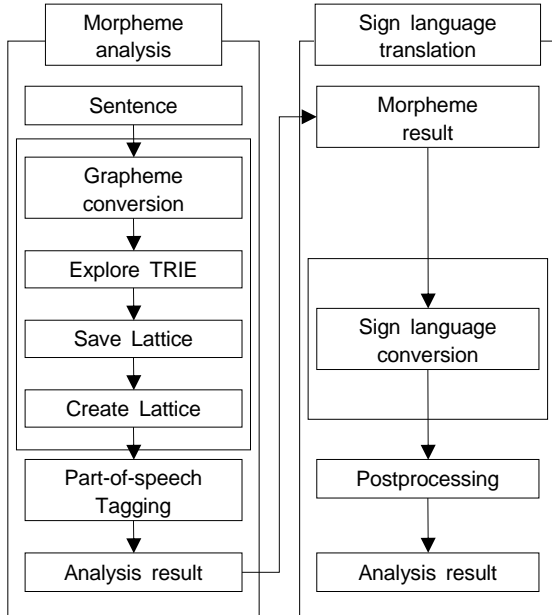


Fig. 1 Morpheme analysis and sign-language translation sequence flow

수화 번역 단계에서 수화 언어 변환이 본 논문에서 제안하는 알고리즘이다. 형태소 분석된 결과를 수화 문장으로 구성할 때 수화의 언어적 특성을 고려한다. 수화 번역 알고리즘은 3단계를 거치게 된다. 각 단계에서 사용한 언어적 특성은 다음과 같다.

### 3.1. 수화 표현을 위한 문장 요소 제거

수화는 한국어에 비해 선어말 어미나 어말 어미, 조사 등 문법적 기능을 제대로 하지 못한다. 문장 성분과 문장 내 단어의 의미 관계를 통해 전체 내용 파악에 문제가 없는 경우 생략한다[3].

- (1) 나는 학교에 간다. → [나+학교+간다]
- (2) 내가 만들었다. → [나+만들다]

(1)의 문장에서 ‘는’, ‘에’와 같은 조사는 ‘나’와 ‘학교’, ‘가다’의 단어 조합에서는 전체 내용 파악에 문제가 없으므로 생략한다. (2)에서 ‘만들었다’는 ‘만들다’라는 형용사와 ‘-었-’이라는 시간적 어미의 결합단어 이

지만 언제 만들었는지는 내용상 중요하지 않기 때문에 생략한다.

### 3.2. 수화 표현의 변환 및 시제 표현

한국어로는 무언가를 셀 때 단위로 표현이 가능하지만 수화로는 따로 표현하는 방법이 없다. 명사+‘하다’로 표현될 때에도 ‘하다’를 한 번 더 표현해주지 않아도 된다.

- (3) 공모전에 신청한 사람이 백 명이 넘는다.  
→ [공모전+신청하다+사람+백+넘다]
- (4) 이전에 새로운 길을 개척하였다.  
→ [이전+새롭다+길+개척하다+끝]

(3)에서 ‘백 명’은 수사+단위로 표현되었다. 앞에 ‘사람’이라는 단어가 단위를 설명할 수 있기 때문에 수사만 표현한다. (4)의 ‘개척하다’는 형태소 분석 과정을 지나면 ‘개척’과 접미사 ‘하다’로 나누어지는데, 수화로 표현할 때 접미사는 따로 표현하는 단어가 없다. 또한 ‘이전’이라는 과거를 뜻하는 시제를 표현하는 부사가 있는 경우 뒤에 ‘끝’이라는 단어가 붙는다. 그러나 ‘끝’과 같은 시제 표현을 하지 않아도 ‘이전’이라는 단어를 통해 과거를 전달하는 점에서는 문제가 없다[5].

### 3.3. 수화 높임말 용어 변경 및 위치 이동

한국어는 부모님이나 연세가 많은 분들에게 높임말을 사용한다. 수화에서는 높임말을 표현하는 단어는 있지만 잘 사용하지 않는다. 또한 수화 문장에는 내용상 강조하고자 하는 부분을 도치하는 표현이 자주 나타난다. 도치하는 방식의 어순 변동은 의문문에서 자주 나타난다[3].

- (5) 너희 아버지 집에 계시니?  
→ [너+아버지+집+있다+?]
- (6) 누가 음식을 훔쳐 먹었을까?  
→ [음식+훔치다+먹다+누구+?]

(5)에서와 같이 ‘계시다’라는 높임말을 ‘있다’와 같이 표현해준다. (6)에서는 ‘누구’라는 단어를 뒤로 도치함으로써 음식을 훔쳐 먹은 사람이 누구인지를 강조한다.

#### IV. 수화 번역 성능평가

앞서 설명한 수화 번역 알고리즘을 사용하여 한국어 문장을 수화 언어로 번역 할 때, 얼마나 정확하게 번역 하는지에 대한 성능 평가가 필요하다. 한국어 문장의 수화 번역 성능이 어느 정도인지의 기준을 나타내는 평가 방식이 존재하지 않기 때문에 본 논문에서는 수화 번역의 성능 평가 기준을 제시한다.

수화 번역에 대한 평가 기준과 점수를 절대 평가 기준과 상대 평가 기준으로 나눈다. 절대 평가 기준에서는 명제내용(15점), 문법구조(25점), 단어인식(25점), 완성도(30점)의 4개 항목이 포함되며, 상대 평가 기준에서는 번역속도(5점)를 포함한다. 성능 평가를 최대

100점으로 하였다. 각 평가 방식별 등급과 설명은 표 1 과 같다.

절대 평가 기준에서 첫 번째 기준인 명제 내용은 문장에 대한 길이와 포함된 단어들의 어려운 정도를 나타낸다. 두 번째 문법구조는 조사나 어미가 알맞게 생략 되었는지, 도치해야 하는 문장이 있는지 등에 대한 내용을 말한다. 세 번째 기준인 단어인식은 문법구조에서 생략되지 못한 단어들이나 대체되어야 하는 단어들이 대체되지 않는 등에 대한 내용을 다룬다. 마지막으로 완성도 평가 기준은 앞서 평가한 명제내용, 문법구조, 단어인식에 대한 평균 점수를 나타낸다. 주로 문법구조와 단어인식 부분이 완성도 점수를 결정한다. 상대 평가 기준인 번역 속도는 형태소 분석과 알고리

Table. 1 Sign-language translation evaluation method

Absolute Evaluation Metrics	(1) Sentence Content(15)	
	Hard (15)	The sentence is long and words are difficult to understand.
	Normal (10)	The sentence is long but words is simple. Or it is short but it has difficult words. (Short sentence: less than 4 words)
	Easy (5)	The sentence is short and easy.
	(2) Grammatical structure (25)	
	Excellent (25)	The sentence is expressed quite grammatically.
	Normal (17)	The grammatical structure has some minor errors.
	Poor (8)	The sentence follows no grammar rules.
	(3) Words(25)	
	Excellent (25)	Each word is translated and its meaning is correct.
	Normal (17)	In short sentence, one or two words is not translated correctly. In long sentence, two or three words is not translated correctly.
	Poor (8)	each word in a sentence is not translated correctly and its meaning is different.
	(4) Completeness(30)	
	Excellent (30)	The sentence is translated grammatically and words are used correctly. The meaning of the sentence is clear.
	Great (24)	The sentence is translated grammatically and some words are used wrong. But the correctness of the sentence is high and it is understandable.
Normal (18)	The sentence is translated normally. The words are used with some errors. The correctness of the sentence is low but it is understandable.	
Poor (12)	The sentence is not translated grammatically. But the meanings of the words are somewhat understandable.	
Very Poor (6)	The sentence is not translated grammatically and the meanings of the words is not understandable.	
Relative Evaluation Metrics	(5) Translation Speed(5)	
	Fast (5)	from 1sec to under 3sec
	Normal (3)	from 3sec to under 5sec
	Slow (1)	from 5sec to under 7sec

음을 통해 걸린 시간을 나눈 것이다. 문장마다 걸리는 시간은 다르지만 6단어 이하의 평균적인 문장은 모두 1초 내외이다.

수화 번역 알고리즘을 거친 100개의 문장을 수화 번역 평가 기준을 통해서 점수화하여 얼마나 정확하게 번역되는지 측정하였다. 평가에 사용한 100개 문장은 국립국어원, 한국농어인협회에서 발행한 한국수화연구제2집에서 다양한 난이도의 문장들로 수집하였다[7]. 표 2는 100개의 문장 중 일부 예시로, 기준 번역은 수화책자에서 제시한 문장 번역 방법으로 제대로 된 수화 번역의 기준이 된다. 제안한 수화 번역 알고리즘을 사용한 결과에 대해서 표 1의 수화 번역 평가 기준 (1)~(5)를 참고하여 주관적으로 등급을 평가하였다.

본 논문에서 제안한 알고리즘으로 완성도를 ‘Great’

이상 받은 문장은 84문장으로, 그 중 완벽하게 번역된 문장은 46문장이다. 나머지 38문장은 굳이 필요하지 않은 단어가 추가되었거나 시제표현이 잘 되지 않았지만 문장의 이해도는 높은 경우이다. 완성도 ‘Normal’ 미만을 받은 문장은 100문장에서 16문장으로 이중에서 형태소 분석 오류로 인한 잘못된 번역이 1개, 시제 및 단어 의미 분석 오류로 인한 잘못된 번역이 15개로 나타났다. 기준 번역과 제안 번역의 결과를 사용하여 본 논문에서 제시한 평가 기준에 따라 각 문장의 수화 번역 결과를 점수화 하였다. 문장의 평가 점수는 본 논문에서 제시한 3단계 알고리즘의 각 단계마다 측정하였다. 평가 결과 1단계 알고리즘만을 거친 후에는 100점 만점에 평균 65.9점이 나왔고, 2단계 과정까지 거친 후에는 79.1점이 되었다. 최종적으로 3단계 알고리즘까지

Table. 2 Evaluation score of proposed sign-language translation

Sentence		나는 너와 서로 좋아하게 되었다.				
Reference translation		나+너+서로+좋아하다+되다				
Suggested translation		나+너+서로+좋아하다+되다				
(1)	(2)	(3)	(4)	(5)	Total	
10	25	25	30	5	95	
Sentence		그는 새로운 땅을 개척하였다.				
Reference translation		그+새롭다+땅+개척하다+끝				
Suggested translation		그+새롭다+땅+개척하다				
(1)	(2)	(3)	(4)	(5)	total	
10	17	25	24	5	81	
Sentence		책상에 있던 연필이 어디론가 사라졌다.				
Reference translation		책상+있다+연필+사라지다+어디				
Suggested translation		책상+있다+연필+사라지다+어디				
(1)	(2)	(3)	(4)	(5)	total	
10	25	25	30	5	95	
Sentence		당신은 진실한 친구를 몇 명이나 가지고 있습니까?				
Reference translation		당신+진실하다+친구+세다+몇				
Suggested translation		당신+진실하다+친구+몇+가지다				
(1)	(2)	(3)	(4)	(5)	total	
10	17	17	18	5	67	
Sentence		왜 학생들이 싸움을 할까.				
Reference translation		학생+싸움+왜				
Suggested translation		학생+싸움+왜				
(1)	(2)	(3)	(4)	(5)	total	
10	25	25	30	5	95	

적용할 경우는 82.1점이 되었다.

## V. 시스템 구현

본 논문에서 제안한 수화 번역 알고리즘을 이용하여 전자책의 한국어 문장을 수화 동영상으로 보여주는 안드로이드 애플리케이션을 구현하였다. 전체적인 애플리케이션의 동작 순서는 그림 2와 같다.

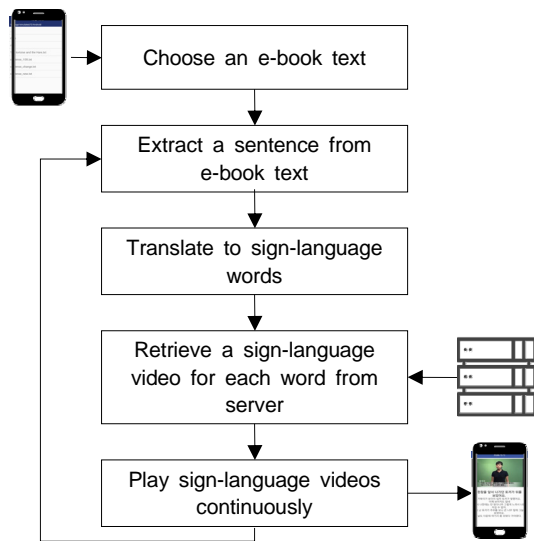


Fig. 2 E-book to sign-language translation application procedure

사용자가 e-book 텍스트 파일을 선택하면 첫 번째 문장을 추출한다. 추출된 문장은 문장 단위로 형태소 분석과 수화 번역 알고리즘을 거쳐 수화 문장으로 번역된다. 번역된 수화 문장에 들어가는 수화 형태소들은 단어와 품사, 서버에 저장되어 있는 동영상의 url이 저장되어 있는 데이터베이스를 통해 매치시킨다. 매치된 동영상을 저장하고 있는 서버에서 가져온다. 본 논문에서 사용한 수화 동영상은 국립국어원의 한국 수어 사전을 참고하였다[8]. 그림 3의 오른쪽 화면과 같이 스마트 디바이스 화면에 수화 동영상과 내용을 출력하여 보여준다. 또한 재생되고 있는 동영상이 어느 문장인지를 사용자가 바로 알 수 있도록 해당 문장은 강조 효과를 주었다.

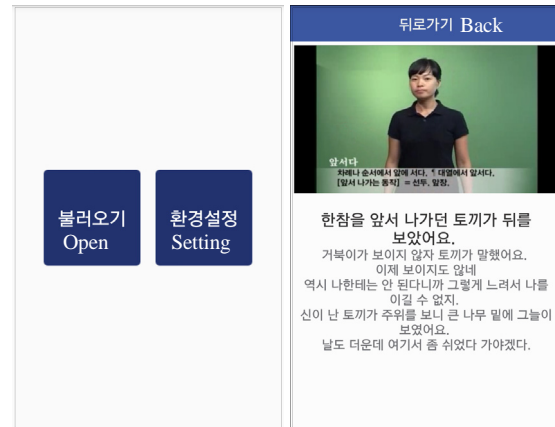


Fig. 3 Implementation of e-book to sign-language translation application

## VI. 결론

본 논문에서는 한국어 문장을 형태소로 분석하여 수화 언어로 번역하는 알고리즘을 제안하고, 수화 번역의 성능을 평가할 수 있는 기준을 제시하였다. 제안한 알고리즘을 이용하여 전자책의 내용을 수화 언어로 번역하여 수화 동영상을 재생하는 안드로이드 기반 애플리케이션을 구현하였다. 이를 통해 청각장애인들이 쉽게 책에 대한 이해도를 높일 수 있을 것이다.

제안한 수화 번역 알고리즘에서는 대부분 단어에 대한 의미 오류로 인해 번역의 완성도가 낮아지므로 문장 내의 단어 의미에 대한 분석을 추가적으로 수행하면 수화 번역 성능을 높일 수 있을 것으로 기대된다.

## REFERENCES

- [1] H. S. Paeng, J. H. Lee and J. Y. Cho, "Research for Sign Language Translation Application Development Base on Smartphone Environment," in *Proceedings of HCI Korea*, pp. 62-65, Jan. 2012.
- [2] S. H. Kim, E. Y. Chang and J. C. Park, "Translating a Complex Sentence in Korean into a Sign Language Script for an Automatic Sign Language Generation," in *Proceeding of the 19th Annual Conference on Human and Cognitive Language Technology*, pp. 15-22, Oct. 2007.

- [3] J. W. Choi and J. Chul Park, "Analysis and Computational Processing of Sentences in Korean for Automatic Sign Language Generation," in *Proceeding of the 15th Annual Conference on Human and Cognitive Language Technology*, pp. 219-226, Oct. 2003.
- [4] H. Y. Jung, E. J. Min, S. J. Ahn, Y. J. Oh and J. H. Lee, "Sense-Tagged Dictionary based Machine Translation from Korean to Korean Sign Language," in *Proceeding of Korea Computer Congress*, pp. 583-585, June 2014.
- [5] J. Y. Kim, J. W. Chung and J. C. Park, "Analyzing and Mapping Expressions of Tense for Korean-Korean Sing Language Translation," in *Proceeding of the 39th KIISE Fall Conference*, pp. 121-123, Nov. 2012.
- [6] Shineware. Komoran 2.0 software [Internet]. Available: <http://www.shineware.co.kr>.
- [7] Korean standard sing-language specification committee, *Korean sign language*, 2nd ed. National Institute of Korean Language, 2010.
- [8] National Institute of Korean Language. Korean Sign Language Dictionary [Internet]. Available: <http://sldict.korean.go.kr>.



한솔이(Sol-Ee Han)

2017년 : 한밭대학교 컴퓨터공학과 졸업(학사)  
 컴퓨터공학과  
 ※관심분야 : 모바일 응용 프로그램



김세아(Se-A Kim)

2017년 : 한밭대학교 컴퓨터공학과 졸업(학사)  
 컴퓨터공학과  
 ※관심분야 : 모바일 응용 프로그램



황경호(Gyung-Ho Hwang)

1998년 KAIST 전기및전자공학과 졸업(공학사)  
 2000년 KAIST 전자전산학과 졸업(공학석사)  
 2005년 KAIST 전자전산학과 졸업(공학박사)  
 2005년 ~ 2007년 삼성전자 무선사업부 책임연구원  
 2007년 ~ 현재 한밭대학교 컴퓨터공학과 부교수  
 ※관심분야 : 이동통신 프로토콜, 무선 센서 네트워크, 모바일 응용 프로그램