

□ 특 집 □

악보 인식 기술 현황

삼성 전자 황 영 길
 육군사관학교 원 영 주*
 서강대학교 김 성 천**

● 목

차 ●

I. 서 론	IV. 인쇄 악보의 인식
II. 영상처리	V. 결 론
III. 패턴 분류 및 인식	

I. 서 론

최근 컴퓨터 기능의 확장 및 통합에 의해 멀티미디어 마인드가 확산되고 있는 추세이며 이를 뒷받침하는 특수 장치의 개발 및 소프트웨어 개발도 큰 진전을 보이고 있다.

정보화 사회의 발전과 함께 컴퓨터를 인간의 기능에 대신하도록 하기 위해서 이를 여러 각도로 발전시켜 왔으며, 이로인해 컴퓨터는 단순한 계산 기능이외에 추론 및 결정 기능과 음성 인식 및 출력기능, 그리고 시각기능 즉 문서 정보의 인식기능을 갖게 되었다.

이러한 제 기능을 구현하기 위해 필요로 하는 기술 중의 일부가 영상 처리와 패턴 분류에 의한 인식기법이다. 이러한 기반위에서 여러가지 형태의 문자와 기호들을 인식하게 된다. 이들 여러 인식 분야 중, 우리는 여기서 패턴 인식의 특수 형태 중의 하나인 악보 인식에 대한 연구를 살펴보기로 하겠다.

문서화된 정보를 인식하기 위해 키보드를 사용하지 아니하고, 컴퓨터의 보조장치인 시각장치를 통해 입력받으며, 이 데이터를 목적에 맞게 적당한 정보로 변환함으로써 입력에 필요한 노력을 최소화하고 입력을 자동화 한다.

이러한 문서 입력의 자동화는 문자인식 분야에서 많은 연구가 이루어져 왔으며 현재에도 여러 방향에서 연구 중에 있다. 인쇄체 문자 및 필기체 문자에 대한 연구는 여러나라에서 각국의 문자 인식에 대한 관심으로 연구가 활발하더 기타 여러가지 도형 및 사물에 대한 연구도 많이 이루어졌으나 인쇄악보의 인식에 대한 연구는 아직 미흡한 상태이다. 인쇄악보의 인식은 악상 기호의 특징을 고려하여 주로 패턴 매칭과 히스토그램(histogram)을 이용하여 이루어진다.

II. 영상 처리

영상 처리는 영상과 영상 데이터를 다루는 과학으로서 폭넓은 응용 분야에 적합한 광범위한 기법을 포함한다[1]. 영상 처리는 다음과 같은

* 정회원
 ** 종신회원

곳에 적용된다. 영상의 외양(appearance)을 강화(highlight)하거나 수정, 또는 그 안에 있는 특정 정보를 강화한다. 그리고 영상내의 요소를 분류, 매칭, 측정한다. 또한 영상의 일부분들을 결합하거나 몇몇 요소를 재구성한다.

영상 처리 알고리즘은 네가지로 분류될 수 있는데 점 처리(point processes), 구역 처리(area processes), 프레임 처리(frame processes), 기하학적 처리(geometric processes)가 그것이다.

점 처리는 영상의 픽셀값을 전적으로 그 픽셀의 값(그리고 때로는 그것의 위치)을 기초로 하여 변경하는 알고리즘이다. 변환에 있어 다른 픽셀값은 관련되지 않는다. 개개의 픽셀은 알고리즘상 그 픽셀의 원래값에 관련되는 새로운 값으로 교체된다. 예로 영상 밝게하기(image brightening), 네가티브 영상(negative image), 영상 스레시홀딩(image thresholding), 영상 콘트라스트 신장(image contrast stretching), 영상 의사색채(image pseudo-coloring) 등이 있다.

그리고 히스토그램(histogram)이란 것은 영상 또는 영상의 일부에 대한 픽셀-강도값의 분포 그래프로서 영상의 전반적인 밝기와 콘트라스트에 관한 많은 것을 나타내고 영상 처리에 있어서 질적, 양적인 정보를 제공한다.

구역 처리는 영상에 관한 정보를 도출하기 위해 픽셀 그룹을 사용하며 예로써 영상 특징 추출(모서리 강화 및 발견), 영상 날카롭게 하기, 영상 부드럽게 하기, 영상 희미하게 하기, 영상의 랜덤 노이즈 제거가 있다.

영상에 포함된 시각적 노이즈를 줄이기 위해서는 low-pass spatial filters를 사용하고, 영상의 모서리를 강조하기 위해서는 high-pass spatial filters를 응용한다. 모서리 강조에는 라플라스 모서리 강조, 이동-차 모서리 강조, 경사 방향 모서리 강조가 있다. 그리고 영상에 포함된 랜덤 노이즈를 제거하기 위해 median filtering을 사용한다.

프레임 처리는 새로운 영상을 산출하기 위해 결합 함수와 함께 두 영상(또는 그 이상의 영상들)으로부터의 정보를 사용한다. 이것은 보안, 품질관리, 영상 질 향상에 응용되기도 하며 비디오 영상의 비교, 압산, 인쇄 회로 기판의 부품

검사에도 사용될 수 있다.

기하학적 처리는 픽셀-강도값을 반드시 변경하는 것은 아니며, 대신 픽셀의 위치를 변경 즉, 주어진 픽셀에 대한 강도값(픽셀에 포함된 정보)이 새로운 위치로 이동된다. 이러한 기하학적 변환 방법에는 scaling, sizing, rotation, transformation, mirror 등이 있다.

III. 패턴 분류 및 인식

패턴이란 벡터 또는 행렬 표기법으로 표현되는 측정치 또는 관측치의 집합이다[2]. 그리고 패턴 인식은 그 측정치의 묘사 또는 분류에 관심을 갖는 과학이며 이것은 종종 인텔리전트 시스템의 중요한 한 요소가 되기도 하고 데이터 전처리와 결정 문제(decision making)에 사용되기도 한다.

패턴 인식은 또한 신호(파형, 영상, 다차원 데이터 등)로부터 정보를 추출하기 위한 기법이며 방법은 크게 두 가지로 분류할 수 있다. 첫째는 통계적 방법(statistical method)과 구조적 방법(structural method)을 포함하는 전통적인 방법(conventional method)이 있으며 두번째는 인공지능에 기반을 둔 방법(AI-based method)이다[3, 8].

통계적(또는 결정론적: decision-theoretic) 패턴 인식 방법은 각 패턴을 feature vector(특성 측정치의 집합)로 표시하고 인식은 feature space를 분리함으로써 이루어진다. 즉 특성함수를 이용한 입력 패턴과 표준 패턴을 비교함으로써 이루어진다.

구조적(또는 구문론적: syntactic) 패턴 인식 방법은 features의 상호관계 또는 상호연결이 중요한 구조적 정보를 제공함으로써 구조적인 묘사 또는 분류를 쉽게 할 수 있다. 즉 패턴을 서브패턴이나 프리미티브로 분해 해나가는데 이는 syntax rule에 의해 패턴 구조를 파싱(parsing)함으로서 이루어진다. 이 방법은 논리적, 체계적이지만 변형에는 부적합하다.

인공 신경망(ANN: Artificial Neural Networks)을 이용한 방법은 인간의 신경망 조직을 모형으로 하여 많은 수의 단순한 프로세스들을 망으로 연결한 시스템을 사용하여 패턴을 인식하

고자 하는 연구로서 신경망 연구의 초기에서부터 진행되어 왔으며, 특히 최근들어 패턴 인식에 대한 많은 연구 결과가 나왔다.

이렇게 패턴 인식에 신경 회로망을 이용하는 것은 종래의 방법론으로는 해결하기에 매우 비효율적인 문제에 대해 아래와 같은 특징을 가진 신경 회로망이 패턴 인식 문제를 효율적으로 해결하는데 매우 적합하리라고 여겨지기 때문이다. 신경 회로망은 패턴 인식에 있어 자주 발생하는 잡음이 있거나 패턴이 변형되고 왜곡된 애매한 데이터를 효과적으로 처리할 수 있다.

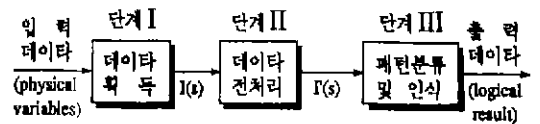
신경 회로망을 이용한 시스템은 학습과 기억을 통하여 문자를 인식하는 동적인 시스템이다. 이러한 동적인 특성으로 인하여 훈련되지 않은 다른 패턴에도 약간의 학습과정을 통하여 쉽게 적응할 수 있는 가능성이 있다. 신경 회로망 시스템은 간단한 학습에 의해 시스템 매개변수 조정이 가능하므로 다른 방식에 비해 개발 기간이 훨씬 단축될 수 있다.

이와같이 신경망을 이용한 패턴 인식은 학습에 의해 새로운 패턴에 적응할 수 있으며, 패턴의 국부적인 변형 및 잡음에 민감하지 않은 장점을 가지는 반면, 패턴의 크기가 큰 경우 학습에 걸리는 시간이 매우 길어지며, 인식 후보 대상의 수가 많은 경우에 성능이 저하되는 단점을 갖는다.

한편 패턴 인식을 위한 시스템은 대개 3단계(그림 1)로 분류할 수 있는데 데이터 획득 단계, 데이터 전처리 단계, 그리고 결정 단계이다[4]. 데이터 획득 단계에서는 실세계(real world)로부터 아날로그 데이터를 변환기(transducer)를 통해 수집한 다음 이를 컴퓨터가 처리하기에 적합한 디지털 형태로 변환한다.

이 결과 데이터는 다시 다음 단계인 전처리 단계의 입력으로 사용되는데 여기에서 몇개의 특성으로 그룹을 형성하여 마지막 단계인 결정 단계에서 각 객체는 분류함수에 의해 분류 즉 인식된다.

패턴 인식의 일부인 문자 인식에 대한 연구는 우리가 쉽게 접할 수 있는 정보가 대개 문자인 이유로 활발하며, 각 국의 문자 특성에 따라 각기 다른 방법으로 자국의 문자 인식에 관심을 갖는



I(t): 측정된 디지털 형태의 데이터, I(s): 분류된 특성함수의 집합
(그림 1) 패턴 인식 3단계.

데 우리나라도 한글 인식에 대한 연구가 있어왔으며[9,10] 연구 현황에 대한 기사가 한국정보과학회지 91년 2월호에 특집으로 실린 바 있다.

IV. 인쇄악보의 인식

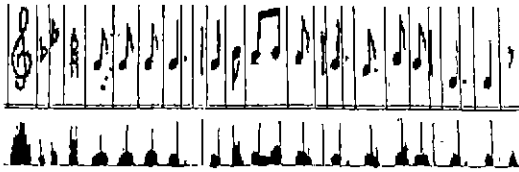
문자 인식이 패턴 인식의 일부인 것과 마찬가지로 악보 인식도 패턴 인식의 특정 분야로서 악상 기호 및 그 위치를 인식하는 것을 목표로 하며, 크게 도형과 문자로 구분할 수 있다. 악상 기호는 악보를 그리는 방법 즉 기보법(또는 정사법-淨寫法)[11]에 따라 정해진 위치에 일정한 크기와 모양으로 도장이나 처인 등을 새겨서 그곳에 잉크를 묻혀서 찍게 된다.

악보 인식에 대한 연구는 문자 인식과 비교하여 아직 미미한 상태이며, 국내외에서 악보를 인식하여 음을 발생시키거나 로보트를 이용하여 연주하는 시스템에 관한 연구가 있다[5,6]. 그리고 속도 향상을 위해 병렬 처리 시스템에서 구현한 사례도 있다[7]. 물론 지금까지의 연구는 인쇄 악보에 제한되어 있으며 인식 가능한 악상 기호의 범위도 각기 다르다.

논문 [5]에서는 인쇄악보 영상을 CCD 카메라를 이용하여 마이크로 컴퓨터에 입력하며 히스토그램(histogram 또는 가산 투영치: integral projection)을 이용하여 각 악상기호들을 분리하고 각 방향에서의 특징 벡터를 구한 결과를 이미 존재하는 기준 패턴과 비교하여 패턴을 분류 즉 악상기호를 인식하게 된다.

이 논문에서는 악보 인식에 가장 기본이 되는 오선지상의 음표(단순음표, 잇단음표, 단순 화음음표, 잇단 화음음표)인식에 한정하고 그 인식된 결과를 연주를 위한 출력코드로 발생하였다.

먼저 영상을 입력하는데 있어서 잡음이 생기는 데 이 잡음을 제거하기 위해 3*3 4-근방 마스크를 사용하여 독립 고립점을 삭제하게 된다.



(그림 2) 오선내에서의 수직 히스토그램.

그리고 각 오선들의 좌표를 추출하기 위해 가로와 세로 방향으로의 히스토그램을 사용하는데 이는 오선 특성에 의해 가능하다.

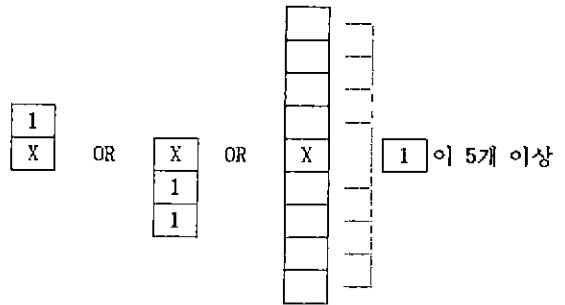
그리고 오선내에서 마디 좌표를 추출하기 위해 오선안을 수직방향으로 가산 투영(그림 2)하여 그 값이 오선간의 간격과 허용된 범위내에서 일치하는 좌표를 찾는다. 마디 추출이 끝나면 마디내의 음표를 인식하기 위해 수직방향으로 가산 투영하여 각 음표의 특징에 근거하여 각각 분류, 인식한다.

인식된 결과, 발생하는 출력 코드를 컴퓨터 내장 스피커를 이용하여 음으로 출력하므로써 연주가 이루어진다. 한 화면내에서 처리할 수 있는 음표의 수는 5-8개에 불과하며 인식률은 약 90% 이상을 기록했다.

논문[6]에서는 카메라로 인쇄 악보를 입력하여 인식결과를 로보트를 이용하여 연주하는 시스템에 관한 연구이다. 인식 알고리즘은 계층적 구조를 가지며 전처리(preprocessing) 단계를 거친 영상을 개략적 분류(coarse classification) 알고리즘에 의해 몇개의 그룹으로 나누고, 이 그룹내에서 다시 세부적인 분류(fine classification)를 하게 된다.

이 연구에서의 전처리 단계는 인식하고자하는 music score image에서 binary image화 하고 오선을 제거하여 인식 단계에서의 인식률을 높이기 위한 것이다. 전처리 단계중 모서리를 명확하게 강조하기 위해 3*3 커널(kernel)을 사용하였고 오선을 찾기 위해서는 히스토그램의 분리성(separability)을 이용하였다.

여러개의 수평성분을 갖는 후보 보표중에서 겹치지 않는 다섯개를 선택하여 각각의 수직 좌표값을 추출하고, 그 간격의 평균값을 보표의 간격으로 하며 상하로 보표 간격의 3배를 설정하여 보표 윈도우를 만든다. 그리고 symbol



(그림 3) 픽셀이 삭제되지 않을 조건 3가지.

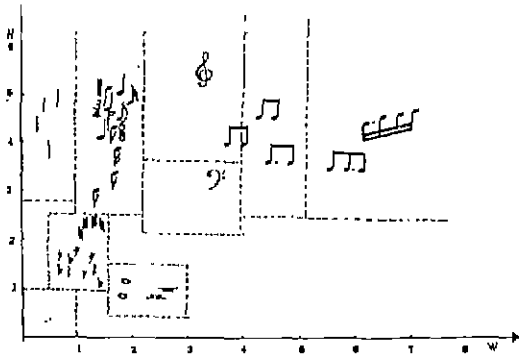
shape recognition을 어렵게 하는 보표를 제거하는데 그 조건은 다음의 세가지 조건(그림 3)을 만족하는 경우가 아닐 때 그 픽셀은 삭제된다.

보표가 삭제된 후 인식을 위한 단계로서 먼저 보표 윈도우내에서 horizontal projection을 하고, 그 다음 각 symbol들을 나누어 주는 symbol window를 만든 후 각각의 symbol에 대해서 개략적 분류를 한다.

개략적 분류의 기준은 horizontal projection을 취한 그것들의 높이와 넓이 및 공백에 의한 특징에 있다. 이러한 projection figure의 특징으로 symbol 전체를 몇개의 그룹으로 나누어 주고 이 그룹내에서 다시 세부 분류를 하게 된다. 개략 분류는 각 symbol을 정규화시켜서 9개의 그룹(그림 4)으로 나누어 진다.

그리고 세부 분류는 music symbol의 특징상 어떤 그룹은 존재 유무만 확인하고 어떤 그룹은 shape만 인식하며, 또 어떤 그룹은 shape와 그것이 존재하는 위치까지 인식하도록 한다. 이를 위해 후보 보표 정보, projection 값에 대한 정보, 개략 분류의 정보 및 symbol 윈도우의 좌표값 등을 이용한다. 그리고 인식된 symbol을 음악으로서의 올바른 기능을 수행할 수 있도록 pitch, duration, symbol의 그룹에 따른 reserved integer value를 갖는 3-D 토큰을 발생시키며 마지막으로 이 결과 데이터를 programmable sound generator와 로보트에 인터페이스하여 인식된 악보를 실제 연주하도록 한다.

이 논문의 인식률은 95~98%를 나타내며 전처리 시간은 약 7분 30초, 그리고 개략 분류에 소요되는 시간은 약 24초 정도이며 세부 분류는



(그림 4) 개략적 분류를 위한 그룹.

실시간으로 처리한다. 그리고 music syntax analysis에 2~3초를 필요로 하고 처리되는 symbol의 수는 100~120개 정도이다.

논문[7]은 특정 출판사에서, 단일 musical font, fixed sized feature로 인쇄된 악보를 읽어 들여서 컴퓨터가 조작가능한 음악 데이터로 변환하는데, 이 과정의 대부분을 병렬컴퓨터(SIMD: Single Instruction-stream Multiple Data-stream)인 Connection Machine(CM)을 사용하여 처리하며 Connection Machine은 각각 8 K byte의 메모리를 갖는 16K 프로세서로 구성되어 있다.

그리고 프로그램을 위한 코드는 Symbolics Common Lisp로 쓰여졌고 lisp machine에서 제공되는 object oriented programming system을 사용하였다. 그리고 Apple scanner를 이용하여 300 dpi에서 읽어들이며 8·1/2*11 inch의 악보 크기가 2544*3300 영상이 된다.

Fixed sized feature(note heads, accidentals, clefs, time signatures)는 항상 그 크기와 모양이 일정하기 때문에 가변적일 때보다 문제가 단순해진다. 즉 원형 정합(template matching)에 의해 이러한 feature를 알 수 있다. fixed sized feature가 아닌 beams, staff, stems, ties, 그리고 slurs들은 크기가 일정하지 않으므로 다른 방법으로 인식하여야 한다.

Beam의 인식은 beam이 note간의 연결이 직선으로 이루어지고 두께가 두꺼우므로 연결성을 이용하여 box를 형성하고 모서리를 찾아서 필요한 매개변수를 계산하기 위해 모서리 점과 같

이 계산을 한다. 그리고 staff lines을 찾는 방법은 horizontal sum으로서 vertical position을 구할 수 있다.

여기에서 staff lines에 의한 peak값과 beams 및 ties에 의한 peak값을 구분하기 어려울 때는 horizontal runs이 길지 않은 부분을 제거(filter-acts like a low pass horizontal filter)하여 ties, beams, 그리고 다른 features에 의한 영향을 감소시키거나, 한 페이지를 여러개의 vertical segment로 분할하여 각 segment의 peak값으로 결정한다.

Measure lines은 thin vertical thing 이외에 제거되고 다음에 horizontal slice로 분할된다. 그리고 각각은 아래 방향으로 값이 누적되어 임계값 이상이면 line segment가 된다. stem을 식별하기 위해 먼저 평균 stem의 두께보다 넓은 horizontal run을 갖는 부분을 제거하고 이로인해 발생하는 staff line에서의 삭제를 복구하기 위해 위, 아래의 point를 검사하여 black이면 복구한다. 그리고 이의 실행은 CM2(16k 프로세서를 갖는) 상에서 페이지당 약 2~3분이 소요된다.

V. 결 론

앞에서 영상 처리에 대한 일반적인 방법과 패턴 인식의 방법론에 대해서 살펴보았으며 또한 패턴 인식의 특정 분야인 인쇄악보의 인식에 대한 몇가지 연구 사례도 살펴보았다.

문서 정보의 입력을 편리하게 하기 위해 온라인 또는 오프라인 문자 인식에 대한 많은 연구와 이의 상용화를 위한 노력이 있어 왔으며 현재 인식 시스템이 상용화되고 있는 추세이다.

국내에서의 악보 인식에 대한 연구는 수년전에 이루어졌으나 지속적인 연구가 이루어지지 못하고 있는것 같다.

그러나 악보 인식에 대한 연구는 아직 미흡한데, 요즈음의 노래문화 확산에 기인하여 악보 인식의 자동화가 요구되므로 좀더 많은 연구가 있어야 할 것이고 또한 이의 실용화를 위해 PC 수준에서의 개발이 필요하며 악상 기호의 인식 알고리즘은 다른 정형적인 도형의 인식과도 결

합하여 응용될 수 있을 것이다.

그리고 논문[7]의 CM2를 이용한 병렬 알고리즘의 구현과 같이 영상 처리의 고속성을 고려한 병렬 처리에 대한 연구도 더 필요할 것 같다.

본고는 악보인식 연구에 대한 개괄적 현황을 기술하였다. 최근 기존의 연구들을 보완하고 인식을 및 인식 범위를 향상시키는 논문이 발표된 바 있으며[9] 현재 대해 테크닉스(대표: 오세영) 주식회사에서 정부의 공업발전 자금으로 PC에 적용하는 상품화를 위한 연구를 수행하고 있다.

참 고 문 헌

1. Craig A. Lindley, "Practical Image Processing: Acquisition, Manipulation and Storage," John Wiley & Sons, Inc., 1991.
2. Robert J. Schalkoff, "Pattern Recognition: Statistical, Structural and Neural Approaches," John Wiley & Sons, Inc., 1992.
3. Zahid Hussain, "Digital Image Processing: Pratical Applications of Parallel Processing Techniques," Ellis Horwood Limited, 1991.
4. Sing-Tze Bow, "Pattern Recognition and Image Preprocessing," Marcel Dekker, Inc., 1992.
5. 이명우, 최중수, "컴퓨터비전 시스템에 의한 인쇄 악보의 인식과 연주," 전자공학회지, 22권 5호, pp. 10-16, 1985.
6. 김완주, "인쇄 악보 인식시스템의 구현에 관한 연구," 한국과학기술원 석사학위 논문, 1986.
7. Alan Ruttenberg, "Optical Reading of Typeset Music," Master of Science in Visual Studies at the MIT, 1991.
8. N. Nandbakumar and J. K. Aggarwal, "The Artificial Intelligence Approach to Pattern Recognition- A Perspective and an Overview," P. R., vol. 18, No. 6, pp.383-389, 1985.
9. 황영길, 최창훈, 김성천, "인쇄 악보의 인식과 병렬

알고리즘에 관한 연구," 한국정보과학회 추계학술대회 논문, 1993.

원 영 주

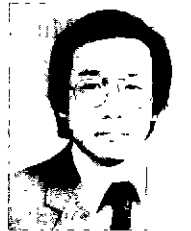


1981 육군사관학교 전자공학과 이학사
 1985미네소타대 집산학 석사
 1987미네소타대 집산학 박사
 1988~1993.6 육군사관학교 수학과 조교수
 1993.6 육군사관학교 수학과 부교수
 관심분야: 알고리즘, 병렬처리 시스템, 멀티미디어



황 영 길

1991경북대 전자공학과 학사
 1993서강대 공대 전자계산학과 석사
 1993~현재 삼성전자 연구원



김 성 천

1975서울대학교 공과대학 공업교육학 (전기전공)학사
 1976~1977 동아컴퓨터(주) Sys. Eng.
 1977~1978 스페리유니백 Sales Rep.
 1979 Wayne State Univ. 컴퓨터공학 석사
 1982 Wayne State Univ. 컴퓨터공학 박사
 1982~1984 캘리포니아주립대 조교수
 1984~1985 금성반도체(주) 책임연구원
 1986~1989 서강대학교 공과대학 전자계산소 부소장
 1989~1991 서강대학교 공과대학 전자계산학과 학과장
 1985~현재 서강대학교 공과대학 전자계산학과 조교수 (1985. 8~1987. 8), 부교수(1987. 9~1992. 8), 교수 (1992. 9~현재)
 1989~현재 한국정보과학회 병렬처리시스템 연구회 부위원장, 대한전자공학회 및 한국통신학회 논문지 편집위원(1991, 1993), 한국정보과학회 학회지 부위원장(1993)
 관심 분야: 병렬처리시스템(Paralledl Computer Architecture, Interconnection Network), Neural Network, Computer Network