

시각정보처리과정을 이용한 인공시각시스템

서창진
상명대학교 정보통신공학과

Artificial Vision System using Human Visual Information Processing

Chang-Jin Seo

Dept. of Information and Telecommunications Engineering, SangMyung University

요약 본 논문은 인간의 생물학적 시각정보처리과정 특성과 웨이블릿을 이용한 인공시각시스템을 제안한다. 인공 시각시스템은 인간의 생물학적 시각정보 처리과정을 이용하며 맹인의 인공시각시스템 제작 및 일반적인 인식시스템에 적용할 수 있다. 본 논문에서는 인간의 망막에서 신경절 세포까지 정보 처리과정을 모델링하여 구성하였고 신경절 세포에서 후두엽 초기시각피질까지 전달되는 정보 복원과정을 모델링하여 초기시각피질에 표현되는 영상정보를 구성하였다. 실험과정은 인간의 시각정보 처리과정 특성인 저주파, 고주파 분해를 웨이블릿 변환으로 시각 정보를 구현하였고 이를 이용하여 인식 시스템에 적용하였다. 실험에 사용한 데이터는 AT&T 얼굴데이터베이스를 사용하였다. 그리고 제안하는 인간의 시각정보처리 과정 특성을 이용한 방법이 영상인식 시스템의 정확성을 향상시킬 수 있음을 실험을 통하여 증명하고자 한다.

주제어 : 인공시각시스템, 웨이블릿, DNN, 얼굴인식, 딥러닝

Abstract In this paper, we propose the artificial vision system using human visual information processing and wavelet. Artificial vision system may be used for the visually impaired person and the machine recognition system. In this paper, we have constructed the information compression process to ganglion cells from the human retina. And we have reconstructed the primary visual information using recovery process to primary visual cortex from ganglion. Primary visual information is constructed by wavelet transformation using a high frequency and low frequency response. In the experiment, we used the faces database of AT&T. And the proposed method was able to improve the accuracy of face recognition considerably. And it was verified through experiments.

Key Words : Artificial Vision, Wavelet, DNN, Face Recognition, Deep Learning

1. 서론

컴퓨터 시각(computer vision)이란 인간의 시각체계와 유사한 인공 시각체계를 만드는 것이라고 말할 수

있다. 컴퓨터를 이용하여 인간의 시각 기능을 대체하는 기술인 컴퓨터 시각은 심리학, 생물학, 기계공학, 의학, 반도체 기술 등과 결합하면서 비약적인 성장의 계기를 마련하고 있다. 알고리즘에서만 가능했던 기법들이

* 본 논문은 2013년 상명대학교 교내연구비에 의하여 지원되었음

Received 17 September 2014, Revised 25 October 2014

Accepted 20 November 2014

Corresponding Author: Chang-Jin Seo(Dept. of Information and Telecommunications Engineering, SangMyung University)

Email: cjseo@smu.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. 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.

반도체기술과 ASIC설계 기술의 성장과 함께 하드웨어 상의 구현이 현실화되었고, 인간의 시각정보 처리과정에 대한 지식들이 밝혀짐에 따라 이를 인공시각에 응용하려는 노력들이 활발하다. 인공시각 기술은 영상처리 알고리즘이나 물체인식, 물체추적 및 자동화제어 방법의 기반 기술이 될 뿐 아니라, 그 자체로도 많은 응용분야를 가진다[1].

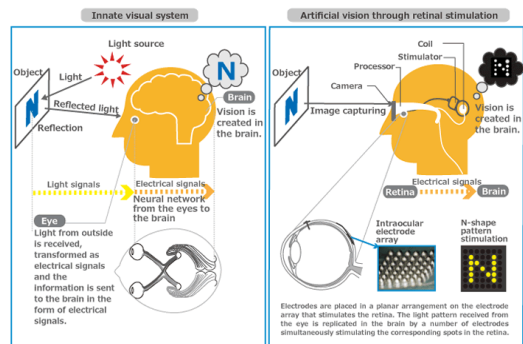
의학적, 심리학적으로 인간의 시각 시스템 계통의 정보처리과정에 대한 지식들이 규명되어 지고 있으며, 이러한 인간의 시각정보 처리과정에 대한 모델링을 통하여 현재 개발되어진 컴퓨터 시각 기술의 성능을 극대화하기 위한 연구뿐만 아니라 인공시각 개발을 위한 연구가 진행되고 있다[2,3].

인간의 시각 계통의 기능 일부를 기계에 제공하는 것을 목적으로 하는 연구가 컴퓨터 시각이며, 그 중에서 영상인식은 2차원 영상의 정보를 추출하여 대상을 인식하는 분야이다. 그리고 인간의 시각계통의 정보 처리과정을 생물학적인 특징 기반으로 개발하는 것이 인공시각(artificial vision)이다.

2. 관련연구

인간의 뇌의 구조를 분석한 시각정보처리과정 분석의 역사는 그리 길지 않다. 인간의 뇌의 구조를 분석한 시각정보 처리과정의 연구는 1979년 David H. Hubel과 Torsten N. Wiesel의 "Brain Mechanisms of Vision"의 연구에서부터 시작된다[4]. 이들의 연구는 초기시각피질의 뉴런이 시각적 형태를 어떻게 부호화 하는지에 대해 수많은 정보를 제공해 주었다. 이 업적으로 두 사람은 1981년 노벨상을 수상했다. 이 연구에서부터 인간의 뇌의 구조를 분석하여, 시각정보 처리과정 모형에 대한 연구가 시작되었지만, 뇌를 분석하는 도구 및 하드웨어 소프트웨어의 성능 부족으로 이 연구는 1990년대 말까지 이론으로만 구현되었으며 기계시각이론의 발전을 가져주었다. 2000년대 부터 fMRI 영상분석 및 생체신호처리 분야에 대한 연구가 활발히 진행되면서, 인간의 시각정보 처리과정이 규명되어지고 있으며 뇌의 정보처리과정에 기반한 시스템을 개발하기 위하여 연구가 진행중에 있다.

인공 시각 시스템에 관련한 또 다른 형태의 연구를 살펴보면 초기시각피질의 특정부위를 자극하면 맹인이 시각정보를 느낀다는 점이다. Wm. H. Dobbelle은 이런 사실에 기반하여 인간의 초기시각피질 영역에 전자회로를 부착하여 맹인에게 실험한 결과, 전체적인 시각은 복원하지 못하였지만 물체의 형태정보(edge)를 인지하게 할 수 있다는 사실을 알게 되었다[5]. 일반적인 인공시각시스템은 [Fig. 1]와 같이 연구되어지고 있다[6].



[Fig. 1] Configuration of an artificial vision system[6]

3. 인간시각 특성과 웨이블릿

인간은 시각적 특징을 분석하고 인식하기 위하여 망막에 맺힌 영상정보를 저주파성분 및 고주파성분으로 영상을 분해한다. 이렇게 주파수 대역으로 영상정보를 분해하는 이유는 망막에 입력되어지는 특징을 추출하기 위해서 이다. 인간의 시각은 빛에 의하여 반응을 하며 망막의 수용기에서 신경질 세포로 정보가 전송될 때 약 80%가 저주파 성분이며, 20%가 고주파성분으로 분해된다고 알려져 있다[7]. 이러한 시각의 정보는 전체 정보에서 20%는 안구의 반사작용(saccadic movement)에 활용되어지고 80%는 물체의 위치 및 형태를 인지하는 과정에 적용되고 있다[8]. 이러한 인간의 시각적 정보의 특성을 분해하기 위해서 영상의 특징을 저주파 및 고주파 대역으로 분해가 필요하다. 웨이블릿 변환은 시간 및 주파수에 대하여 지역성을 가지는 신호로 표현할 수 있기 때문에 비정체성 신호를 가지는 영상을 해석하는 데 유리한 점을 가지고 있다. 웨이블릿으로 분해된 영상은 인간 시

각과 유사한 주파수 대역을 가지고 있는 것으로 분석되고 있다. 인간 시각에 민감한 부분에 비하여 상대적으로 둔감한 영역의 영상정보를 많이 제거함으로써 영상인식에 필요한 효율적인 영상을 생성할 수 있다. 최근 들어 인간 시각의 특성을 고려한 연구들이 활발히 진행되고 있다. 웨이블릿 변환은 신호의 비정상성 신호(non-stationary signals)나 신호의 경향성(trends), 신호의 단절성(breakdown points), 유사성(self similarity), 다중해상도분석(multi resolution analysis), 정보의 저주파, 고주파 분해능과 같은 신호의 과도기적 특성을 분석하는 능력을 가지고 있다. 이러한 웨이블릿 변환의 특성은 인간의 시각정보처리과정의 특성과 유사하다[9].

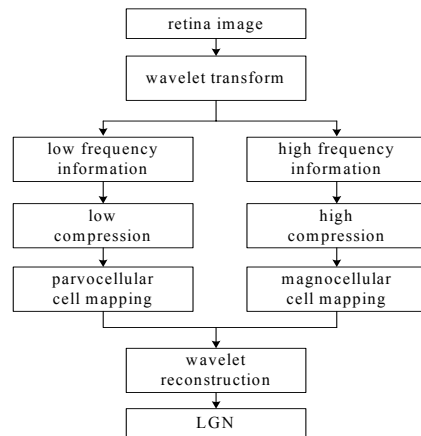
4 제안하는 연구방법

4.1 망막에서 외측슬상핵(LGN)모형

인간의 망막구조는 매우 복잡한 구조로 구성되어 있다. 인간의 망막은 약 1억3천만개의 수용기로 구성되어 있다. 망막은 빛의 신호를 신경 신호로 변환을 수행하는 곳이며, 약 1억3천만의 수용기(receptor) 정보는 약100만개의 신경절 세포로 전송되는 과정에서 물체의 형태정보 및 특징정보를 분류하여 축약되는 과정을 수행하고 있다. 이때 망막에서 신경절 세포로의 전송되어지는 정보는 130:1의 정보 압축을 수행하고 있으며 물체를 인식하는데 필요한 중심와(fovea)의 정보는 축약정보의 80%정도로 물체형태의 인식과정을 담당하는 신경절 세포의 Parvo-cell로 매핑이 이루어진다. 또한 물체의 위치정보를 인식하는데 필요한 정보는 축약정보의 20% 정도로 물체의 위치를 검출과정을 담당하는 신경절 세포의 Magno-cell로 매핑이 이루어진다[10]. 영상 정보의 압축은 매우 엄청나게 이루어지지만 인간은 저주파, 고주파 성분 분해능을 가지며 인식에 필요한 경로 따라 다양한 데이터의 특성 분류를 하는 시각정보 처리특성을 가지므로 인지과정에서 별 다른 문제없이 물체를 인식하고 있다.

본 논문은 인간의 저주파, 고주파 분해 정보 표현을 위하여 인간의 시각 특성을 잘 표현하는 EZW(embedded zerotree wavelet)를 이용하여 망막에서 외측슬상핵까지의 정보압축과정을 모델화 하였다. [Fig. 2]에서 인간의

망막에서 외측슬상핵(LGN) 세포까지의 정보축약과정을 나타내고 있다.



[Fig. 2] The schematic diagram from retina to LGN

4.2 시각정보분해 및 축약

망막에 맺히는 영상은 빛감지세포(photo-receptor)에 의하여 빛의 신호를 전기화학적 신호로 변화가 이루어지며 신경절 세포로 정보 전송이 이루어진다. 이때 신경절 세포는 초기시각피질로 정보전송을 위하여 영상의 신호를 저주파 성분과 고주파 성분으로 분해를 하며 영상정보를 축약하는 기능을 수행한다. 이때 망막에서 신경절 세포로의 매핑과정은 물체를 인식하기 위한 경로(what-pathway)와 물체의 위치를 인식하기 위한 경로(where-pathway)를 담당하는 신경절 세포의 Parvo cellular cell과 Magno cellular cell로 정보의 분화가 이루어진다. Parvo cellular cell은 일반적으로 물체를 인식하기 위한 경로이며 p-pathway라 불리기도 하며 일반적으로 영상의 색상 및 저주파의 성분으로 구성되어져 있다. Magno cellular cell은 일반적으로 물체의 위치를 인식하기 위한 경로이며 m-pathway라 불리기도 하고 일반적으로 영상의 윤곽정보 및 고주파의 성분으로 구성되어져 있다.

본 논문에서는 이러한 인간의 시각적 특성을 웨이블릿의 저주파 및 고주파성분 분해능을 이용하여 모형화한다. 영상의 저주파 성분은 아래의 식(1)을 이용하여 추출하였다. 그리고 영상의 고주파 성분은 아래의 식(2)을 이용하여 추출하였다.

i) Smoothed Signal(저주파 필터 계수에 의해 구해지는 부분)

$$c_k^{J-1} = \sum_{m=0}^{N-1} h(m)c_{M+2k}^J \quad (\text{eq 1})$$

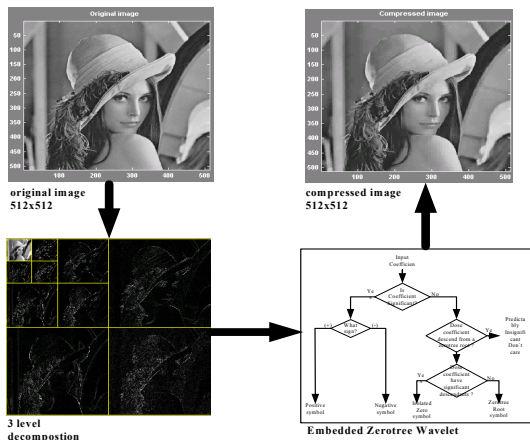
ii) Wavelet Detail(고주파 필터 계수에 의해 구해지는 부분)

$$d_k^{J-1} = \sum_{m=0}^{N-1} g(m)c_{M+2k}^J \quad (\text{eq 2})$$

인간 안구의 중심와(fovea)영역의 중요한 데이터는 무엇(what)에 관련한 형태정보이므로 손실을 최대한 억제하며, 중심와 주변(peripheral)영역의 데이터는 어디(when)에 관련한 위치 정보이므로 최대한 작은 정보로 수렴하게 되어있다.

시각정보 처리 축약과정에서 방향성분 $O=HH$ 일 때, 대각 방향의 에지 성분을 가지는 고주파 성분에 대하여 $1/\sqrt{2}$ 의 가중치를 부여한다. 수평, 수직 성분보다 대각 성분이 대략적으로 $\sqrt{2}$ 의 크기만큼 인간 시각에 둔감하게 반응한다는 것으로 실험적으로 밝혀져 있다[11]. 본 논문에서는 이러한 대각 성분의 둔감성을 이용하여 영상축약과정에 적용하였다.

4.3 EZW(Embedded Zerotree Wavelet)



[Fig. 3] Compression processing of the image

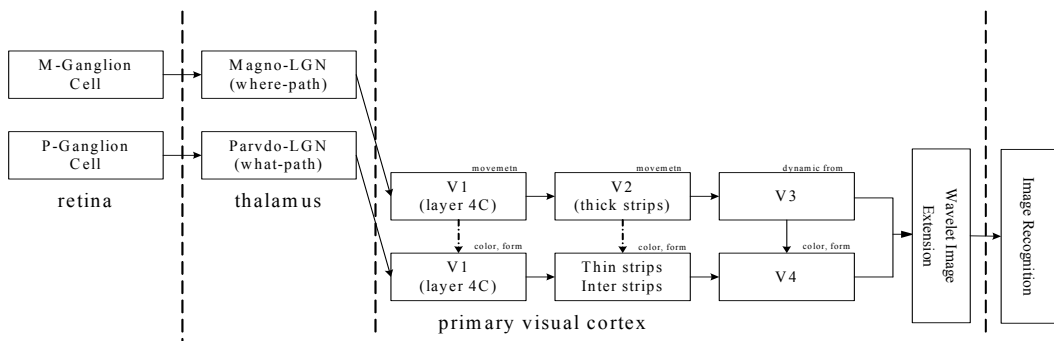
EZW는 웨이블릿 계수를 부호화 하기 위하여 Shapiro가 제안한 방법으로 높은 압축율을 가지면서도 알고리즘이 간단하고, 중요한 정보가 비트 스트림의 앞부분에 오도록 하는 embedded의 특성을 지니고 있다[9]. EZW에서는 웨이블릿 변환 계수들을 중요한 계수와 중요하지 않은 계수로 나누어서 중요한 계수들의 위치에 대한 정보와 중요한 계수의 값에 대한 정보로 나누어서 전송하는 방법을 이용한다.

중요한 계수들의 위치를 나타내는 중요도 맵을 압축 전송하는 단계는 dominant pass라 한다. 중요하다고 판단된 계수는 subordinate pass라고 한다. 본 논문에서는 이러한 중요도 계수를 이용하여 인간시각정보를 압축 표현하였다[Fig. 3]. 웨이블릿 계수는 주어진 역치값 T 를 기준으로 계수 x 가 조건 $|x| \geq T$ 를 만족하면 중요한 계수(significant coefficient)로 그렇지 않으면 중요하지 않은 계수(insignificant coefficient)로 정해진다. 본 논문에서는 최근 웨이블릿 변환 방식중 정지 영상 압축에서 좋은 성능을 보이는 EZW부호화 방법을 사용하여 시각정보를 압축하였다. 식(3)은 초기 역치값의 선정 방법이다.

$$T_{init} = 2^{\lceil \log_2(\max(Y)) \rceil} \quad (\text{eq 3})$$

4.4 LGN 복원

인간의 망막의 정보는 신경절세포로 정보의 전송 과정을 거치면서 최대의 정보축약과정을 거친다. 이후 망막의 특징 지도라 불리는 초기시각피질로 이동되기 이전에 양안의 정보를 외측슬상핵(LGN)으로 전송한다. 이때 외측슬상핵에서는 망막의 특정부위의 자극과 동일하게 반응을 하므로 압축되어진 계수의 표현이 아닌 복원 영상의 표현이다. 이러한 특성은 웨이블릿변환의 특성과 매우 유사하다. 웨이블릿 변환시 저대역의 정보는 원 영상을 표현하는 섬세한 정보를 많이 가지고 있으며, 원영상의 정보량도 많이 존재한다. 그리고 망막의 중심와 영역에서 신경절세포의 P-cell로의 매핑과정과 매커니즘이 매우 유사하다. 그리고, 고대역의 정보는 원영상의 형태 정보를 많이 가지고 있으며 영상의 정보가 저대역에 비하여 다소 적으며, 중심와 주변영역에서 신경절세포의 M-cell의 매핑과정과 매커니즘이 유사하다. 이러한 시각정보의 시각정보 압축 처리과정을 웨이블릿변환을 이용



[Fig. 4] Mapping process from LGN to primary visual cortex

하여 구현하였다.

웨이블릿의 FIR 필터로 h_0 저주파 필터와 h_1 고주파 필터로 두면, 식 (4)과 같이 복원필터를 구할 수 있다.

$$g_0(n) = (-1)^n h_{1(n-1)} \quad (\text{eq 4})$$

$$g_1(n) = (-1)^{n-1} h_{0(n-1)}$$

대역분할은 3 level mallat tree를 이용하였으며, 고대역은 98%가 대역분할을 차지하므로, 부호화하기 위하여 각 대역에 따른 수평, 수직, 대각방향을 고려하여 지그재그 스캔(Zigzag Scan)하여 부호화 하였다. 부호화된 영상을 다시 복호화과정과 역양자화과정을 거쳐서 각 단계 별의 압축된 영상을 구하여 초기시각피질로 전송하기 위한 영상을 구하였다.

4.5 초기시각피질 영상복원

시각정보처리과정이 규명되기 이전에는 망막의 정보를 뇌에 직접 전달하여 인식하는 것으로 알고 있었으나, 최근 fMRI 및 생체신호 분석에 의하면 인간은 망막의 레벨에서 인식에 필요한 정보는 최소 손실압축을 하게 되며 위치를 인식하기 위한 정보는 최대 손실압축을 수행한다. 즉, 전체 영상정보를 뇌의 초기시각피질로 전송하는 것이 아니라 일정한 형태로 영상을 세분화하며 인식에 필요한 정보를 전송하는 것이다. 망막의 영상은 망막에서 신경절세포로 영상 축약과정을 수행한 후 시신경 교차를 통하여 외측슬상핵으로 정보가 전송되어진다. 하지만 초기시각피질로 영상이 전송되어진 정보는 물체의 인지 과정 이전에 망막에서 맺힌 정보량의 약 두배로 확

대 되어진다. 즉, 영상전송에 필요한 정보는 최소화 하면서 영상인식과정에 필요한 영상은 최대한 확장되어진다. 외측슬상핵에서 초기시각피질로의 매핑과정은 [Fig. 4]와 같이 이루어진다.

5. 실험 및 결과분석

본 논문에서는 인간의 시각특성인 저주파 민감성, 고주파 둔감성의 특징을 가지고 망막에서 신경절 세포까지의 영상정보 축약과정과 초기시각피질까지의 영상복원과정을 구성하였다. 그리고 처리된 정보의 인식성능을 확인하기 위하여 입력 패턴에 대하여 정확한 해답을 주지 않아도 생물학적 인식시스템과 유사하고, 스스로 학습이 가능한 비지도학습 알고리즘인 Grossberg의 ART2 알고리즘을 채택하였다[12].



[Fig. 5] AT&T Face Database

<Table 1>의 실험 결과는 높은 압축율에서도 원영상으로 복원하기 위한 정보가 많이 남아있다는 사실을 알

수 있으며 초기시각피질까지의 영상복원과정에서도 원 영상에 근접한 영상표현이 될 수 있다는 사실을 보여주고 있다. <Table 2>는 AT&T 얼굴 데이터베이스[그림 5]를 이용한 시각정보 축약에 대한 인식을 변화 실험결과를 나타내고 있으며 <Table 3>은 최종 시각정보복원영상에 대한 인식을 변화에 대한 실험 결과이다. 실험결과 제안하는 방법으로 구성된 영상정보가 일반적인 영상 압축방법에 비하여 효율적인 인식성능을 가짐을 알 수 있었다.

<Table 1> Image compression using EZW(256×256)

bit/pixel	compression	MSE	PSNR
6	0.75	0.9040	48.5689
4	0.50	0.9040	48.5689
2	0.25	5.7965	40.4992
1	0.125	19.0545	35.3308
0.5	0.0625	49.1714	31.2137
0.0625	0.0078	276.2008	23.7186

<Table 2> Change in the recognition rate for visual information compression

bpp	DCT		Proposed method	
	number of success	recognition rate	number of success	recognition rate
6	155	92.26%	155	92.26%
4	154	91.66%	155	92.26%
2	149	88.69%	155	92.26%
1	148	88.09%	154	91.66%
0.5	143	85.11%	154	91.66%
0.25	130	77.38%	152	90.47%

<Table 3> Change in the recognition rate for visual information reconstruction

bpp	DCT		Proposed method	
	number of success	recognition rate	number of success	recognition rate
6	158	93.45%	161	95.83%
4	155	92.26%	161	95.83%
2	153	91.07%	161	95.83%
1	148	88.09%	157	93.45%
0.5	145	86.30%	155	92.26%
0.25	142	84.52%	153	91.07%

6. 결론

본 연구는 인간의 생물학적 시각정보 처리과정의 특징인 시각정보의 축약과 복원과정을 이용하여 영상인식 모형을 구현하였다. 인간의 시각시스템은 영상정보 전송 시 최대한 압축 전송을 하며 인식과정에 들어가기 전에 영상인식에 필요한 정보를 복원하는 과정을 가지고 있다. 즉, 망막의 영상을 초기시각피질로 모두 전송을 하지 않고 필요한 데이터의 형태로 영상을 전송하여 인식한다는 것이다. 실험결과 영상의 저주파 성분 보다 고주파 성분을 삭제하고 물체를 인식하여도 충분하다는 사실을 알 수 있었다. 실험에 사용한 영상크기는 실험용 영상 DB의 크기 문제로 128x128, 256x256의 사이즈의 영상을 사용하였다. 실제 안구에 맺히는 영상크기에 비하여 상당히 작은 크기이므로 향후 연구에서는 고해상도 영상에 대한 실험이 필요할 것으로 판단한다.

ACKNOWLEDGMENTS

This work was supported by the SangMyung University Research Grants in 2013.

REFERENCES

- [1] Yoshua Bengio, Aaron Courville, Pascal Vincent, Representation Learning : A Review and New Perspectives, Arxiv, 2012.
- [2] Guenther T, Lovell NH, Suaning GJ, 2012 Bionic vision: system architectures: a review. Expert Rev. Med. Devices 9, 33-48, 2012 (doi:10.1586/erd.11.58)
- [3] Martin fisxhler, Osker Firschein, Intelligence: The eye, the brain and the computer, Addison-Wesley, 1987.
- [4] David H, Hubel, Torsten N. Wiese, Brain Mechanisms of Vision, Neuro-Vision, IEEE Press, 1979.
- [5] Wm. H. Dobbelle, Artificial Vision for the Blind by Connecting a Television Camera to the Visual

- Cortex, ASAIO journal, pp. 3-9, 2000.
- [6] Terasawa, Y. Tashiro, H. Nakano, Y. Osawa, K.; Ozawa, M., Safety Assessment of Semichronic Suprachoroidal Electrical Stimulation to Rabbit Retina, Conference Proceedings of the IEEE Engineering in Medicine and Biology Society, 3567-3570, Jul. 5, 2013.
- [7] Rauzettel-Klosinski S. Rehabilitation of lesions in the visual pathways. Klin. Monbl. Augenheilkd. 226, 897 - 907. 2009. (doi:10.1055/s-0028-1109874)
- [8] Eickenscheidt M, Jenkner M, Thewes R, Fromherz P, Zeck G. 2012 Electrical stimulation of retinal neurons in epiretinal and subretinal configuration using a multicapacitor array. J. Neurophysiol. 107, 2742-2755, 2012. (doi:10.1152/jn.00909.2011)
- [9] J. M. Shapiro, Embedded Image coding using zerotrees of wavelet coefficients, IEEE Trans. on Signal Procossing, 41, 12, 3445-3462, Dec. 1993.
- [10] Stingl K, et al. Safety, efficacy of subretinal visual implants in humans : methodological aspects. 4-13. 2013. (doi:10.1111/j.1444-0938.2012.00816.x)
- [11] S. Mallat, Multi-Frequency Channel Decomposition of Images Wavelets Models, IEEE Trans. on Information Theory, vol. 11, no. 7, July 1992.
- [12] Simon Haykin, Neural Networks : A Comprehensive Foundation, MacMillan, 1994.

서 창 진(Seo, Chang Jin)



- 1999년 2월 : 부산대학교 멀티미디어 이학석사
- 2003년 8월 : 부산대학교 멀티미디어 공학박사
- 2000년 3월 ~ 2013년 2월 : 성덕대학교 교수
- 2013년 3월 ~ 현재 : 상명대학교 정보통신공학과 조교수

- 관심분야 : Artificial Vision, Multimedia, E-Learning, Security
- E-Mail : cjseo@smu.ac.kr