

결합된 파티클 필터에 기반한 강인한 3차원 손 추적

(Robust 3D Hand Tracking based on a Coupled Particle Filter)

안 우 석 [†] 석 흥 일 ^{**} 이 성 환 ^{***}
 (Woo-Seok Ahn) (Heung-Il Suk) (Seong-Whan Lee)

요 약 손 추적 기술은 인간과 기계와의 효율적인 의사소통을 위한 손동작 인식 기술의 핵심 기반 기술이다. 최근의 손 추적 연구는 3차원 손 모델을 이용한 연구 방향에 초점을 맞추고 있고, 기존의 2차원 손 모델을 이용한 방법보다 강인한 추적 성능을 보이고 있다. 본 논문에서는 결합된 파티클 필터에 기반한 새로운 3차원 손 추적 방법을 제안한다. 이는 전역적 손 형상과 지역적 손가락 움직임을 분리하여 추정하고, 각각의 추정 결과를 서로의 사전 정보로 이용하여 기존의 방법보다 빠르고 강인한 추적을 가능하게 한다. 또한, 추적 성능 향상을 위해 색상과 에지를 함께 고려한 다중 증거 결합 방법을 적용한다. 실험 결과, 제안하는 방법은 복잡한 배경이나 동작에서도 강인한 추적 결과를 보였다.

키워드 : 결합된 파티클 필터, 3차원 손 추적, 다중 증거 결합, 3차원 손 모델

Abstract Tracking hands is an essential technique for hand gesture recognition which is an efficient way in Human Computer Interaction (HCI). Recently, many researchers have focused on hands tracking using a 3D hand model and showed robust tracking results compared to using 2D hand models. In this paper, we propose a novel 3D hand tracking method based on a coupled particle filter. This provides robust and fast tracking results by estimating each part of global hand poses and local finger motions separately and then utilizing the estimated results as a prior for each other. Furthermore, in order to improve the robustness, we apply a multi-cue based method by integrating a color-based area matching method and an edge-based distance matching method. In our experiments, the proposed method showed robust tracking results for complex hand motions in a cluttered background.

Key words : Coupled particle filter, 3D hand tracking, multi-cue integration, 3D hand model

1. 서론

최근, 3차원 손 모델을 이용한 추적 방법의 연구가 활발히 진행되고 있으며, 강인한 추적 성능을 보이고 있다. 그러나 손은 많은 수의 관절로 이루어져 있고, 자유도가 높아 추적에 필요한 연산량이 매우 많을 뿐 아니라, 손을 대표할 특징들이 적기 때문에 추적이 어렵다. 이를 해결하기 위해 B. Stenger 등은 3차원 손 모델이 가질 수 있는 모든 형태를 미리 얻고 계층적으로 분류하는 방법을 제안했다[1]. 하지만 연산 시간이 오래 걸리고, 모든 가능한 손의 형태를 처리할 수 없다는 한계를 가진다. M. Bray 등은 3차원 추적에 있어서의 고차원 문제를 해결하기 위해 스마트 파티클 필터를 제안했다[2]. 하지만 이는 구조광 센서를 이용한 경우에만 적용 가능하고, 조명 변화와 손과 얼굴이 겹칠 때 쉽게 추적을 실패한다. J. McCormick과 M. Isard는 분할 샘플

· 이 논문은 2009년도 교육과학기술부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2009-0060113). 이 연구에 참여한 연구자는 '2 단계 BK21사업'의 지원을 받았다

[†] 정 회 원 : 고려대학교 컴퓨터학과
wsahn@image.korea.ac.kr

^{**} 학생회원 : 고려대학교 컴퓨터학과
hisuk@image.korea.ac.kr

^{***} 통신회원 : 고려대학교 정보통신대학 교수
swlee@image.korea.ac.kr
(Corresponding author)

논문접수 : 2009년 8월 13일
심사완료 : 2009년 10월 12일

Copyright©2010 한국정보과학회 : 개인 목적이거나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

링 방법을 손 추적에 이용하였으며[3], Y. Wu 등은 샘플링을 이용한 손 추적 과정에서의 많은 계산량을 극복하기 위해 분할 정복 방법을 제안하였다[4]. 이들은 전역적 손 형상을 먼저 추정한 뒤, 그 위치로부터 지역적 손가락 움직임을 추정하기 위해 계층적인 방법을 이용하였다. 전역적 손 형상은 2차원의 손바닥을 추적함으로써 추정할 수 있다고 가정하였다. 그러나 손바닥은 추적에 강인한 특징을 제공하지 못하고, 가려짐 등으로 인해 추적에 실패하는 경우가 많이 발생한다.

본 논문에서는 결합된 파티클 필터 기반의 3차원 손 추적 방법을 제안한다. 결합된 파티클 필터는 전역적 손 형상과 지역적 손가락 움직임을 독립적으로 추정하고, 각각의 추정 결과를 서로의 사전 정보로 이용하는 유기적인 방법이다. 이를 위해 전역적 손 형상과 지역적 손가락 움직임을 각각 형태 기반의 추적에 강인하다고 알려진 파티클 필터[5]를 사용하여 추정하고, 손의 해부학적인 연결 관계를 고려하여 파티클들을 결합한다. 제안하는 방법은 추적에 필요한 연산량을 줄이면서 강인한 추적이 가능하다. 또한, 다양한 환경 변화에서의 추적 성능 향상을 위해 색상 기반의 영역 매칭과 예지 기반의 거리 매칭 방법을 결합한 다중 증거 결합(multi-cue integration) 방법을 적용한다.

논문의 구성은 다음과 같다. 2절에서는 결합된 파티클 필터 기반의 추적 방법 및 다중 증거 결합 기반 매칭 방법에 대해 설명한다. 3절에서는 실험 및 결과를 분석하고, 4절에서는 결론을 맺는다.

2 결합된 파티클 필터 기반의 추적

본 논문에서는 강인한 3차원 손 추적을 위하여, 전역적 손 형상과 지역적 손가락 움직임 정보를 결합한다. 이 과정에서 3차원 손 모델을 사영하여 얻은 2차원 모

델로 실제 영상과 다중 증거 결합 방법을 이용하여 매칭한다.

2.1 3차원 손 모델

다양한 시점에서의 손 형태를 얻기 위해 그림 1(a)와 같은 3차원 손 모델을 이용한다. 임의의 손 형태는 손가락 관절의 자유도를 고려하여 다음과 같이 26차원 벡터로 표현될 수 있다.

$$\begin{aligned}
 X &= [X^G, X^L]^T \\
 X^G &= [x^t, y^t, z^t, x^r, y^r, z^r] \\
 X^L &= [thumb, index, middle, ring, little] \\
 thumb &= [th_{mp1}, th_{mp2}, th_{em}, th_{ip}] \\
 index &= [in_{mp1}, in_{mp2}, in_{mp3}, in_{pip}] \\
 middle &= [mi_{mp1}, mi_{mp2}, mi_{mp3}, mi_{pip}] \\
 ring &= [ri_{mp1}, ri_{mp2}, ri_{mp3}, ri_{pip}] \\
 little &= [li_{mp1}, li_{mp2}, li_{mp3}, li_{pip}]
 \end{aligned} \tag{1}$$

손 X 는 3차원 공간에서의 이동(x^t, y^t, z^t)과 회전(x^r, y^r, z^r)으로 구성된 전역적 손 형상 X^G 와 각 손가락 관절의 움직임 가능한 각도[6]들로 구성된 지역적 손가락 움직임 X^L 로 표현된다. 이 벡터를 기반으로 15개의 원통과 19개의 구를 이용하여 3차원 손 모델을 표현한 것이 그림 1(b)에 나타나 있다. 또한 실제 영상과 3차원 모델 유사도를 비교하기 위해 재구성된 3차원 손 모델을 수직 사영하여 그림 1(c)와 같은 2차원의 손 형태 모델을 만든다.

2.2 추론

기존의 계층적 추정 방법은 그림 2(a)와 같이 전역적 손 형상 $X_t^G = \{x_1^G, \dots, x_t^G\}$ 을 추정한 결과로부터 지역 손가락 움직임 $X_t^L = \{x_1^L, \dots, x_t^L\}$ 을 추정한다[3,4]. 따라서 먼저 수행하는 전역적 손 형상 추정의 결과가 전체 손 추적 결과에 끼치는 영향이 매우 크다.

전역적 손 형상의 추적에 이전 시간 $t-1$ 에 추정된 지

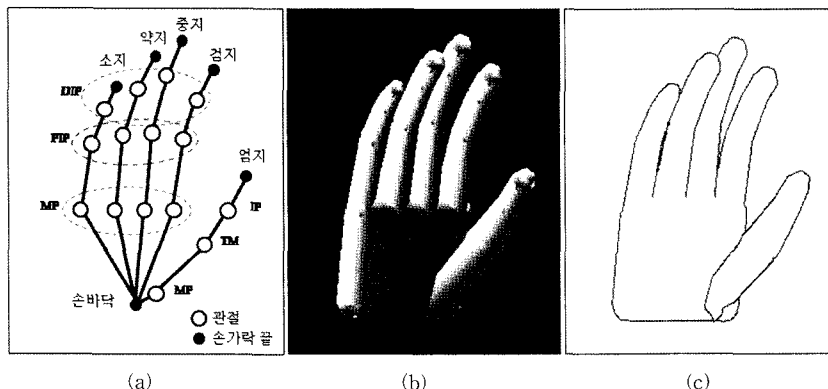


그림 1 3차원 손 모델: (a) 관절로 이어진 손 모델, (b) 재구성된 3차원 손 모델, (c) 수직 사영한 2차원 손 형태 모델

역적 손가락 움직임 정보를 이용한다면 전역 추적 결과에만 의존하는 기존 방법에서의 문제를 해결할 수 있다. 또한 이 방법은 손 전체의 형상 정보를 이용할 수 있기 때문에 손바닥의 형상 정보만을 이용하는 계층적 방법보다 강한 추정이 가능하다. 지역적 손가락 움직임 또한 이전의 방법처럼 전역적 손 형상의 현재 시간에 추정된 값을 사전 정보로 이용한다. 전역적 손 형상의 추정과 지역적 손가락 움직임의 추정에 각각의 파티클 필터를 이용하여 추정하지만 상호보완적으로 정보를 주고 받을 수 있도록 결합된 파티클 필터 모델을 제안한다. 그림 2(b)는 제안하는 결합된 파티클 필터 모델에 대한 그래프 표현을 보여준다.

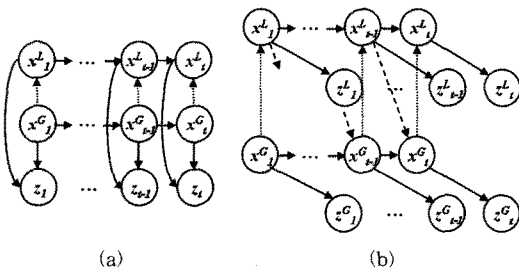


그림 2 손 추적 모델의 그래프 표현: (a) 계층적인 방법 [3,4], (b) 제안하는 방법

전역적 손 형상은 지역적 손가락 움직임의 이전 시간에 추정된 결과를 사전 정보로 이용하여 식 (2)와 같이 표현되고, 주변 확률을 이용하여 추정한다.

$$p(x_i^G | Z_i) = \int_{x_{i-1}^G} p(x_i^G | Z_i, x_{i-1}^G) \quad (2)$$

식 (2)에서 Z 는 관측값들의 정보 $Z_i = \{z_1, \dots, z_i\}$ 이다. 이를 파티클 필터의 샘플링 기법을 이용한 사후 확률 추정식에 적용하면 다음과 같다.

$$\int_{x_{i-1}^G} p(x_i^G | Z_i^G, x_{i-1}^G) = \sum_{n=1}^{N/10} \sum_{m=1}^M p(z_i^{(m)} | g_i^{(m)}, l_{i-1}^{(n)}) p(g_i^{(m)} | g_{i-1}^{(m)}, l_{i-1}^{(n)}) p(g_{i-1}^{(m)} | z_{i-1}^{(m)}, l_{i-2}^{(n)}) \quad (3)$$

여기서 N 은 지역적 손가락의 움직임 추정, M 은 전역적 손 형상의 추정에 사용된 샘플의 수, g 는 전역적 손 형상, l 은 지역적 손가락의 움직임 추정에 사용된 샘플을 의미한다. 샘플링을 이용한 추정 연산에서의 높은 계산량은 Y. Wu 등[4]이 제안한 분할 샘플링 기법을 이용하여 해결한다. 분할 샘플링 기법은 연결된 두 개의 파티클 필터가 사후 확률을 다른 파티클 필터의 사전 확률로 전파할 때 상위 10%의 샘플들로도 충분하다는 것을 이용한다[3]. 따라서 식 (3)의 N 을 $N/10$ 으로 줄일

수 있다.

지역적 손가락 움직임도 현재 추정된 전역적 손 형상 x_i^G 을 사전 정보로 이용한다. 지역적 손가락 움직임의 추정과 파티클 필터의 샘플 계산 방법은 다음과 같다.

$$p(x_i^L | Z_i) = \int_{x_i^G} p(x_i^L | Z_i, x_i^G) \quad (4)$$

$$\int_{x_i^G} p(x_i^L | Z_i^L, x_i^G) =$$

$$\sum_{m=1}^{M/10} \sum_{n=1}^N p(z_i^{(n)} | l_i^{(n)}, g_i^{(m)}) p(l_i^{(n)} | l_{i-1}^{(n)}, g_i^{(m)}) p(l_{i-1}^{(n)} | z_{i-1}^{(n)}, g_{i-1}^{(m)}) \quad (5)$$

전역적 손 형상을 찾은 후, 그 확률 분포에 기반하여 지역적 손가락 움직임을 찾기 때문에 계산 복잡도를 줄일 수 있다. 전역적 손 형상의 추정에도 동일한 방법을 적용한다.

2.3 다중 증거 결합(multi-cue integration) 기반 매칭

손 추적의 성능은 확률 추정에 사용된 관측값과 매칭 방법에 크게 영향을 받는다. 형태 기반의 추적에는 예지 [7]와 색상 정보[8]를 이용한 매칭 방법이 주를 이루고 있다. 특히 피부색 정보를 이용한 특징의 추출과 매칭 방법은 빠르고, 정확한 반면에, 얼굴에 가려지거나, 반팔 옷을 착용한 경우에는 추적에 실패하는 경우가 많이 발생한다. 예지를 이용한 거리 매칭 방법 또한 정확한 매칭이 가능하지만 손과 배경의 대비가 높지 않거나 손의 움직임이 빠른 경우 예지가 추출되지 않는 문제가 있다. 본 논문에서는 예지 기반의 거리 매칭 방법과 피부색 기반의 영역 매칭 방법을 $z = (z_{dist}, z_{area})^T$ 로 결합하고, 다음과 같이 근사화한다.

$$p(z|x) = p(z_{dist}, z_{area}) \approx \alpha p(z_{dist}|x) + (1-\alpha)p(z_{area}|x) \quad (6)$$

여기서 α 는 0에서 1사이의 값으로, 피부색 매칭과 예지 매칭의 의존도를 결정하는 상수이다. 본 논문에서 α 는 실험적으로 얻은 $\alpha = 0.46$ 를 사용했다. 피부색과 예지 기반의 매칭은 각각 Chamfer 거리[9]와 피부색 확률과 형태 모델 기반의 영역 매칭[10]을 이용한다.

3. 실험 및 결과 분석

3.1 실험 환경

실험에 사용된 데이터는 CMOS 카메라를 이용하여 320×240 크기의 24비트 컬러, 초당 25프레임의 영상을 일반 사무실 환경에서 획득했다. 펜티엄 2.4GHz의 PC 환경에서 C++와 OpenCV[11]를 이용하여 구현하였고, OpenGL[12]을 이용하여 가상 데이터와 3차원 손 모델을 생성하였다.

3.2 실험 결과

첫 번째 실험에서는 손이 x, y, z 축으로 이동하며 각 손가락의 움직임이 있는 데이터를 사용했다. 총 214 프레임 길이의 영상을 전역 추정에 사용되는 샘플 수 $M=400$, 지역 추정에 샘플 수 $N=300$ 으로 설정하였다. 제안하는 방법과 Y. Wu 등의 방법[4]을 이용한 비교 실험에서 각 관절의 평균 오차는 각각 3.52° 와 3.56° 로 유사한 성능을 보였다.

두 번째 실험은 손이 x, y, z 축으로 이동과 회전, 각 손가락의 움직임이 있는 데이터를 이용하여 비교 실험을 하였다. 각 $M=500, N=200$ 의 샘플수를 이용하여 추적한 결과, 그림 3에 보여진 바와 같이 Y. Wu 등의 방법[4]은 42번째 프레임부터 추적에 실패한 반면에 제안하는 방법은 전역적 손 형상 추정에 있어서 정확도가 높아 강인한 추적 결과를 보였다.

마지막으로, 다중 증거 결합 기반 매칭 방법의 효율성을 검증하기 위하여 복잡한 실 환경에서 얼굴과 손이 겹칠 때에 대한 실험을 수행하였다. 결합된 파티클 필터에 기반한 추적에서 색상 기반의 매칭 방법, 에지 기반의 매칭 방법, 다중 증거 결합 기반 매칭 방법을 각각

적용하였을 때의 결과를 비교하였다. 그림 4는 각 매칭 방법에 대한 실험 결과이며, y 축은 자승평균 평방근 오차(RMSE)로 다음과 같이 계산한다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n \|x_g(i) - x_c(i)\|^2}{n}} \quad (7)$$

여기서 n 은 프레임 번호이며, x_g 는 실제 손끝의 좌표, x_c 는 추정된 좌표를 각각 의미한다. 손과 얼굴이 겹칠 때 피부색을 이용한 매칭 방법은 신뢰할 수 없고, 매우 복잡한 환경에는 에지 기반과 색상 기반의 매칭 방법 모두 높은 오차를 보인 반면, 제안하는 다중 증거 결합 기반 매칭 방법은 어떠한 상황에서도 10회소 미만의 오차로 강인한 추적 결과를 보였다.

4. 결론

본 논문에서는 결합된 파티클 필터에 기반한 3차원 손 추적 방법을 제안하였다. 제안한 방법은 기존의 계층적인 방법보다 전역적 손 형상 추정에 강인하고, 그로 인해 손 추적 성능이 우수함을 보였다. 또한, 손 추적기

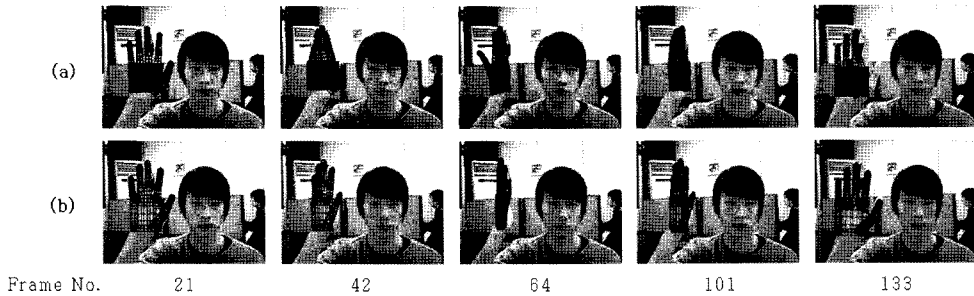


그림 3 실 환경에서의 손 추적 결과 비교: (a) Y. Wu 등의 방법[4], (b) 제안하는 방법

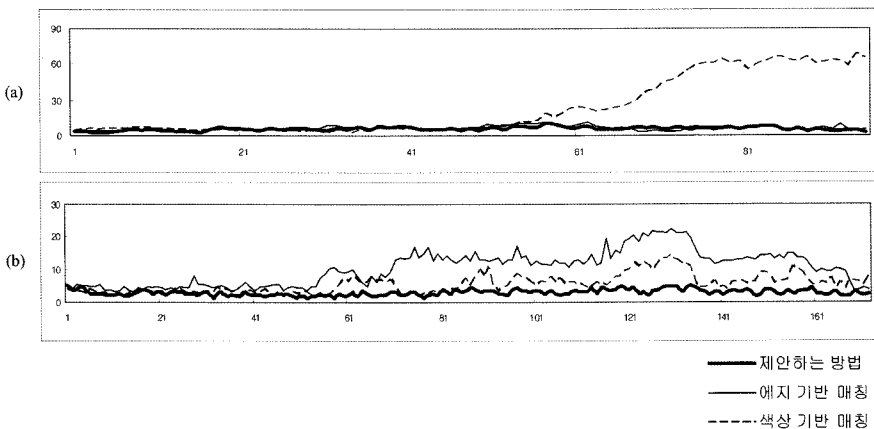


그림 4 추적 결과 비교: (a) 손과 얼굴이 겹치는 상황, (b) 복잡한 환경

의 성능 향상을 위한 다중 증거 결합 방법이 다양한 환경에서 보다 정확한 추적이 가능함을 보였다.

참고 문헌

- [1] B. Stenger, A. Thayananthan, P. Torr, and R. Cipolla, "Model-based Hand Tracking using a Hierarchical Bayesian Filter," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.28, no.9, pp.1372-1384, 2006.
- [2] M. Bray, E. Meier, and L. Gool, "Smart Particle Filtering for High-Dimensional Tracking," *Computer Vision and Image Understanding*, vol.106, no.1, pp.116-129, 2007.
- [3] J. Maccormick and M. Isard, "Partitioned Sampling, Articulated Objects, and Interface Quality Hand Tracking," *Proc. European Conference on Computer Vision, Dublin, Ireland*, pp.3-19, 2000.
- [4] Y. Wu, J. Lin, and T. Huang, "Analyzing and Capturing Articulated Hand Motion in Image Sequences," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.27, no.12, pp.1910-1922, 2005.
- [5] M. Isard and A. Blake, "CONDENSATION - Conditional Density Propagation for Visual Tracking," *International Journal of Computer Vision*, vol.29, no.1, pp.5-28, 1998.
- [6] J. Lin, Y. Wu, and T. Huang, "Modeling the Constraints of Human Hand Motion," *Proc. Workshop on Human Motion, Texas, USA*, pp.121-126, 2000.
- [7] F. Chen, C. Fu, and C. Huang, "Hand Gesture Recognition Using a Real-time Tracking Method and Hidden Markov Models," *Image and Vision Computing*, vol.21, no.8, pp.745-758, 2003.
- [8] A. Argyros and M. Lourakis, "Real Time Tracking of Multiple Skin Colored Objects with a Possibly Moving Camera," *European Conference on Computer Vision, Prague, Czech Republic*, pp. 368-379, 2004.
- [9] G. Borgefors, "Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.10, no.6, pp.849-865, 1988.
- [10] J. Kovac, P. Peer, and F. Solina, "Human Skin Color Clustering for Face Detection," *Proc. The IEEE Region 8 EUROCON 2003: Computer as a Tool*, pp.144-148, 2003.
- [11] <http://sourceforge.net/projects/opencvlibrary/>
- [12] <http://www.opengl.org/resources/libraries/>



안 우 석

2007년 서경대학교 컴퓨터학과(학사)
2009년 고려대학교 컴퓨터학과(석사). 2009년~현재 KIST 연구원. 관심분야는 패턴인식, 로봇 비전 등



석 홍 일

2004년 부경대학교 멀티미디어공학과(학사). 2007년 부경대학교 컴퓨터공학과(석사). 2008년~현재 고려대학교 컴퓨터학과 박사과정. 관심분야는 패턴인식, 뇌-컴퓨터 인터페이스, 컴퓨터 시각 등



이 성 환

1984년 서울대학교 계산통계학과(학사)
1986년 KAIST 전산학과(석사). 1989년 KAIST 전산학과(박사). 1989년~1995년 충북대학교 컴퓨터학과 조교수. 2001년~2002년 MIT Brain and Cognitive Sciences 학과 방문교수. 1995년~현재 고려대학교 정보통신대학 정교수. 2009년~현재 고려대학교 현대·기아차 석좌교수, IEEE Fellow. 관심분야는 패턴인식, 컴퓨터 시각, 뇌공학 등