물체 인식을 위한 질의 기반 주목 알고리즘

류광근, 이상훈, 서일홍
한양대학교

Query-based Attention Algorithm for Object Recognition

Gwang-geun Ryu, Sanghoon Lee, Il Hong Suh
Hanyang University
e-mail : {ggryu, shlee}@incorl.hanyang.ac.kr, ihsuh@hanyang.ac.kr

요 약

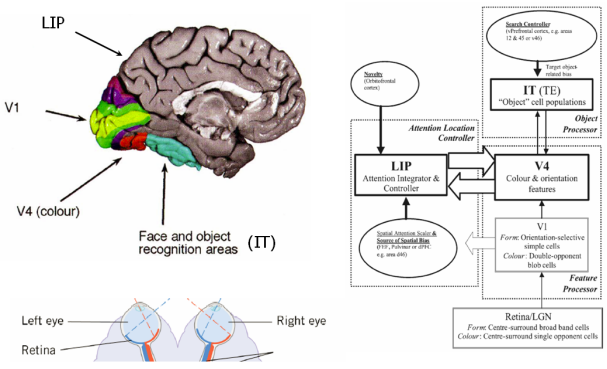
특징점 기반의 물체인식 알고리즘은 물체 인식률을 높이기 위해서 물체에 일정 각도와 거리 이내로 접근해야 한다. 영상처리에서의 주목 알고리즘은 영상의 Low-level Feature들을 조합하여 자극도를 계산한다. 그리고 주목도가 높은 순서대로 시점 이동을 가능하게 한다. 따라서 원하는 물체가 있는 곳의 주목도를 높이면 로봇이 물체 인식을 할 수 있도록 이동 방향을 정할 수 있다. 기존의 주목 알고리즘은 영상의 Low-level Feature를 동등한 비율로 사용하였지만 본 논문에서는 질의를 한 물체를 찾기 위해 질의 물체의 속성을 분석하여 주목 알고리즘을 구성하는 Feature의 크기에 가중치를 줌으로써 질의를 한 물체가 존재할 확률이 높은 순서로 자극도를 강화시키는 방법을 제안한다.

1. 서론

지능로봇은 비전센서를 이용하여 영상으로부터 Semantic정보를 취득할 수 있다. 물체인식은 대표적인 Semantic정보이다. 예를 들면 David. Lowe의 Scale-invariant Feature Transform (SIFT)[1]은 영상에서 강인한 특징점을 추출하여 Database와 비교한 뒤 어떤 물체인지 알아낸다. 이 알고리즘은 물체 인식률이 매우 뛰어나지만 인식률을 높이기 위해서 물체 영상을 얻기 위한 각도와 거리 제한이 존재한다. 따라서 물체 인식률을 높이기 로봇이 목표에 접근하려면 현재상태에서 얻어진 영상 정보로 로봇이 진행할 방향을 정할 수 있어야 한다. 이를 위하여 인간의 주목 알고리즘을 이용한다.

인간은 어떤 물건을 찾을 때 눈으로부터 들어오는 자극과 사전 기억을 이용하여 주목을 해야 할 지점을 찾아낼 수 있다[2]. [그림 1]은 인간의 주목 알고리즘모델과 그에 해당하는 두뇌의 부위를 나타낸다. 하단에서 상단 방향이 눈으로부터 들어오는 자극의 방향이다. 이와 마찬가지로 컴퓨터를 이용한 영상처리에서도 Low-level Feature들을 추출하여 일련의 과정을 거치면 자극도를 계산할 수 있다. 대표적으로 Itti와 Koch의 시각 주목(Visual Attention)에 관한 연구가 있다.[3][4].

영상의 자극도를 계산하는 Bottom-up Approach는 물체가 다수 존재하면 자극도가 무작위로 분포하여 목표 물체를 찾는데 악영향을 줄 수 있다. 따라서 물체가 두개 이상 존재하는 영상에서는 목표 물체의 자극도를 강화하면 정확하게 목표에 주목할 수 있다. 인간의 두뇌에는 사전 기억이나 지식을 기반으로 주목할 곳에 대한 Bias를 주는 IT (Inferior Temporal)라는 기관이 존재한다[2]. [그림 1]의 상단에 위치하는 IT는 Retina로부터 들어오는 시각 자극을 처리하여 인간이 주목할 부분에 대한 자극만을 강화시킨다. 본 논문에서는 IT를 시각 주목 알고리즘에 적용한다. 목표 물체의 Low-level Feature정보



[그림 1] 인간의 두뇌 및 주목 모델

를 계산하여 시각 주목 알고리즘에서 자극도를 구성하는 성분 중에서 목표 물체에 해당하는 성분을 강화시켜 물체가 있는 위치의 자극도를 높이고자 하는 것이다.

영상을 구성하는 특징은 크게 Texture와 Shape로 나눌 수 있다. Texture는 색을 예로 들 수 있고, Shape은 외곽선이 대표적인 예이다. 색은 멀리서도 인지가 가능하지만 외곽선의 경우는 시점에 구애받을 수 있다. 실제로 Low-level Feature들의 시점 이동에 대한 기여도를 조사한 논문[5]에 따르면 색이 시점이 이동하는 위치를 결정하는데 70%이상의 영향을 끼치게 된다. [5]의 연구를 토대로 질의 물체의 색상 성분을 분석하여 자극도 계산을 위한 색상 성분 중에서 질의 물체를 구성하는 색상 순으로 가중치를 부여한다.

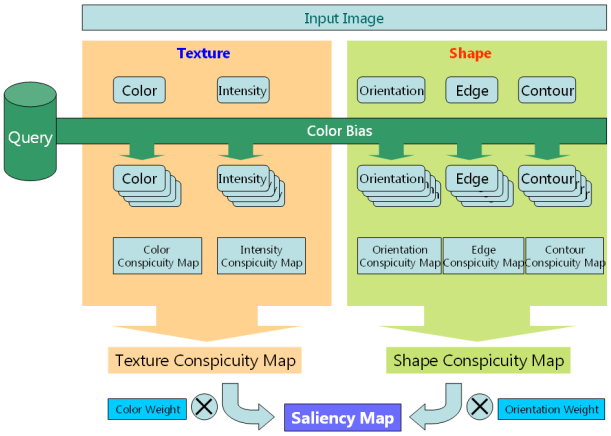
자극도를 나타내는 영상을 구성하는 두드러짐 영상(Conspicuity Map)은 Texture속성의 두드러짐 영상과 Shape기반의 두르러짐 영상의 두 종류가 있다. [5]의 연구를 바탕으로 기존의 1대1이었던 Texture와 Shape의 비율을 7대3으로 변경한다. 즉, 질의 기반 주목 알고리즘을 구현하기 위해 2번의 가중치를 부여한다. 이는 자극도에 색상 성분의 영향을 크게 만들기 위해서이다.

질의 기반 시각 주목 알고리즘에 대한 설명을 2장에서 언급하고 3장에서는 본 논문이 제안하는 알고리즘을 증명하기 위한 실험을 수행한다.

2. 질의 기반 시각 주목 알고리즘

Itti와 Koch의 Bottom-up Attention Algorithm[3][4]은 Low-level Feature를 기반으로 영상의 자극도를 계산한다. Low-level Feature는 색, 밝기, 방향성 등이 있으며 필요에 따라 외곽선, 모션에 관한 성분을 더 추가할 수 있다. 본 논문에서는 색과 밝기 그리고 방향성을 이용한다. 그리고 같은 물체에 대한 시점이동의 반복을 방지하기 위해 외곽선과 Contour정보를 이용한다. 시스템의 구조를 [그림 2]에 나타내었다.

[그림 2]의 좌측에 있는 Query는 질의를 한 물체를 나타내며 이 물체가 각 Feature들에게 영향을 주는 것을 화살표로 나타내었다. 그리고 Saliency Map을 구성하는 Conspicuity Map에 색상과 방향성을 재구성하여 기존의 1대1의 비율을 7대 3으로 변화시켰다.



[그림 2] 시스템 구조도

2.1 입력 영상 및 가우시안 피라미드 구성

색의 종류는 적색(R), 청색(B), 녹색(G), 황색(Y)이다. 그리고 밝기(I)와 방향성(O)를 구한다. 각 정보를 얻어내기 위해 다음의 식을 이용한다.

$$I = (r+g+b)/3 \quad (1)$$
$$R = r-(g+b)/2 \quad (2)$$
$$G = g-(r+b)/2 \quad (3)$$
$$B = b-(r+g)/2 \quad (4)$$
$$Y = (r+g)/2-|r-g|/2-b \quad (5)$$
$$O(0^\circ, 45^\circ, 90^\circ, 135^\circ) \quad (6)$$

식(6)은 Gabor filter[6]를 나타낸다. 총 4개 방향에 대한 방향성을 계산한다. Gabor filter를 계산하는 식은 다음과 같다.

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-\frac{\|k_{\mu,\nu}\|^2\|z\|^2}{2\sigma^2}} [e^{ik_{\mu,\nu}z} - e^{-\frac{\sigma^2}{2}}] \quad (7)$$

식(7)에서 μ와 ν는 Gabor kernel의 orientation과 scale을 나타낸다. $z=(x,y)$이며 $\|\cdot\|$는 norm operator를 나타낸다. $k_{\mu,\nu}=k_v e^{i\phi_\mu}$이며 wave vector를 나타낸다. $k_v = k_{\max}/f^v$, $\phi_\mu = \pi\mu/8$이고 $k_{\max}$는 maximum frequency를 의미한다[6].

식(1)-(6)을 통해 얻은 영상은 시각 주목 알고리즘의 기본 입력이 된다. 이 입력 영상들을 9개의 Scale을 갖는 가우시안 피라미드(Gaussian Pyramid)로 구성한다. 가우시안 피라미드는 다음과 같이 표현한다.

$$I(\sigma), R(\sigma), G(\sigma), B(\sigma), Y(\sigma), O(\sigma,\theta), E(\sigma) \quad (8)$$

인수 σ는 $\sigma \in [0..8]$를 의미하고 θ는

$\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$를 나타낸다.

2.2 질의 물체 및 가중치

영상에서 어떤 물체가 있는지 찾기 위해서는 영상의 일반적인 특징을 얻을 필요가 있다. 가령, 황색병을 찾으라고 할 경우 색에 대한 정보가 들어있는 Low-level Feature에서 황색의 정보를 강화시키고 외곽선에 대한 정보가 들어있는 곳에서 병 모양의 정보를 강화시키면 황색병에 대해서 주목을 할 수 있는 것이다. 질의 물체의 색 정보를 추출하기 위해 HSV 컬러 모델을 이용하였다. 질의 물체의 R, G, B, Y색의 비중을 계산하고 이를 식 (2), (3), (4), (5)의 결과에 반영한다.

2.3 특징 영상 (Feature Maps)

가우시안 피라미드를 이용하여 특징 영상을 계산한다. 어떤 픽셀이 주변에 비해 두드러지는지 찾아내기 위해 Center Surround Difference(Θ) 연산을 이용한다. 특징 영상은 다음의 식을 이용하여 구할 수 있다.

$$I(c,s) = |I(c) \Theta I(s)| \quad (9)$$

$$RG(c,s) = |(R(c) - G(c)) \Theta (R(s) - G(s)) \quad (10)$$

$$BY(c,s) = |(B(c) - Y(c)) \Theta (B(s) - Y(s)) \quad (11)$$

$$O(c,s,\theta) = |O(c,\theta) \Theta O(s,\theta)| \quad (12)$$

인수 c와 s의 의미는 각각 다음과 같다.

$$c \in \{2,3,4\}, s = c + \delta, \delta \in \{3,4\}$$

2.4 두드러짐 영상 (Conspicuity Maps)

특징 영상에 비선형 정규화(N)[7]를 적용한다. 비선형 정규화를 이용하면 영상에서 두드러지는 부분을 강화시키고 그 외의 부분은 억제할 수 있다. 그 뒤에 각 특징 영상을 가우시안 피라미드의 Center Scale과 같은 면적으로 Resize하여 픽셀 대 픽셀 합을 구한다. 이를 Across-scale Addition($\oplus$)이라 한다. 다음 식들을 이용하여 두드러짐 영상을 구할 수 있다.

$$\overline{I} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N(I(c,s)) \quad (13)$$

$$\overline{C} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} [N(RG(c,s)) + N(BY(c,s))] \quad (14)$$

$$\overline{O} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} N\left(\bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N(O(c,s,\theta)) \right) \quad (15)$$

2.5 자극 영상 (Saliency Map)

두드러짐 영상을 모두 합하여 다음 식과 같이 자극 영상을 구축한다.

$$S = \frac{1}{3}(N(\overline{I}) + N(\overline{C}) + N(\overline{O})) \quad (16)$$

식(16)에 색과 방향성의 기여도 비중에 따라 가중치를 부여한다. 실험에 의해 7:3의 비율로 조절하였으며 조절된 자극 영상은 다음 식과 같다.

$$S' = (0.7) \cdot N(\overline{C}) + (0.3) \cdot N(\overline{O}) \quad (17)$$

자극 영상은 입력받은 영상의 자극도 분포를 표현한 것으로 값이 클수록 자극이 높다. 자극이 높은 순서대로 주목을 시작하게 되는데 이렇게 주목하는 시점의 이동을 표현한 것을 Scan Path라 한다.

자극 영상의 자극도의 크기에 따라 시점의 위치를 정하고 일정 면적에 해당하는 Mask영상을 만든다. Mask영상은 자극도 주변 일정 영역은 1의 값을 갖고 나머지 영역은 0의 값을 갖는다. Mask와 입력영상(I)은 다음 식을 이용하여 합칠 수 있다.

$$I'(x,y) = [255 - M(x,y) \cdot (255 - I(x,y))] \quad (18)$$

Mask의 분포에 따라 잘라낸 영상(I')과 질의 영상을 비교하기 위해 SIFT Algorithm[1]을 사용하였다. SIFT의 특징점들을 매치하여 매치된 특징점이 3개 이상일 경우 해당 물체가 인식된 것으로 한다.

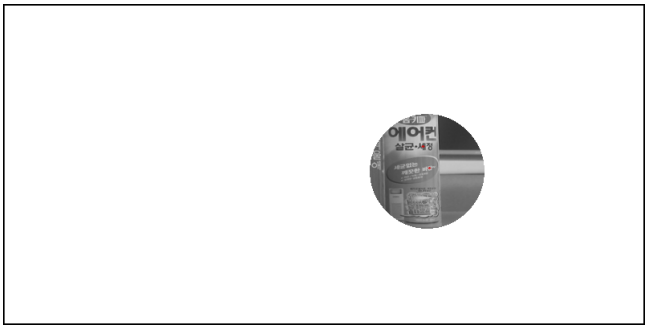
3. 실험 및 결과

실험구성은 다음과 같다. 800x400x24bit의 컬러 영상을 이용하고 질의 물체는 영상 안에 존재하는 물체 들 중 무작위로 선정하였다. 실험의 목적은 물체가 2개 이상 존재하는 Complex Scene에서 질의 물체를 찾는 것이므로 영상 안에는 적어도 5개 이상의 물체를 배치했다. 실험은 다음 4가지 방법으로 진행하였다. 기존 Attention Algorithm을 이용하여 자극도를 계산한 경우, 기존 방법에 색과 방향성의 7:3비

율을 적용하였을 경우, 질의에 따른 색의 가중치 변화를 적용한 경우, 마지막으로 질의에 따른 색의 가중치 변화와 색과 방향성의 7:3비율을 모두 적용한 경우 4가지에 대한 실험을 하였다. 성능의 비교방법은 질의한 물체를 찾을 때까지 움직인 시점의 횟수이다. 즉, 질의에 해당하는 물체를 찾는 횟수가 보통의 주목 알고리즘을 사용하였을 경우보다 적을 경우

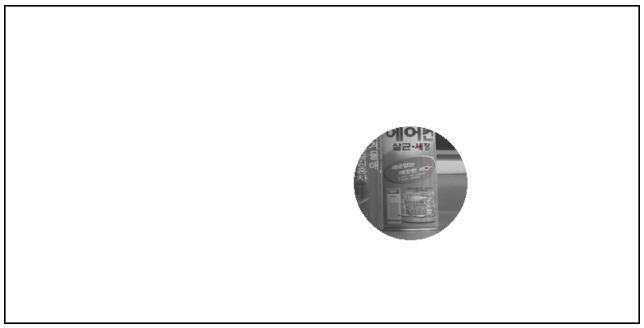
[그림 3] 입력 영상 및 Scan Path 예시
(800x400x24bit)

[그림 4] 질의 영상 예시



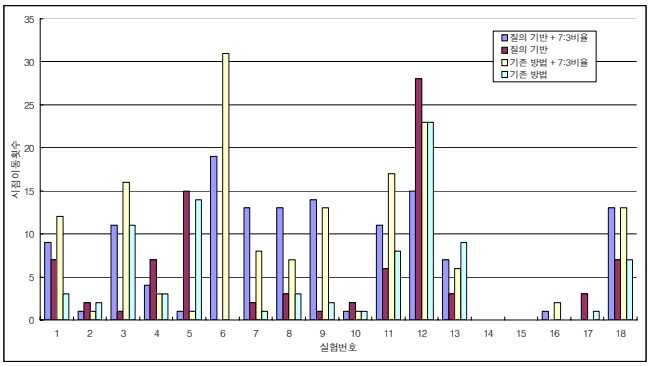
[그림 6] [그림 5]에 주목을 적용한 영상 예시
- 시점 이동 2회 (800x400x8bit)

[그림 7] 질의에 의해 자극도가 변한 영상 예시
(50x25x8bit)



[그림 8] [그림 7]에 주목을 적용한 영상 예시
- 시점 이동 1회 (800x400x8bit)

[그림 5] 기존 알고리즘으로 얻은 자극 영상 예시
(50x25x8bit)



[그림 9] 실험 결과 그래프

[표 1] 실험 결과

	기존 방법	기존 방법 + 7:3 비율	질의 기반 방법	질의 기반 방법 + 7:3비율
평균 시점이동 횟수(검색성공)	6.69	10.27	**6.46**	8.06
3회 이내 검색 성공 확률	44.4%	27.8%	**50%**	27.8%
검색에 실패한 확률	27.8%	22.2%	22.2%	**16.7%**

자극도가 올바로 강화되었다고 할 수 있다. Complex Scene은 5장을 사용하였고 한 Scene당 4번의 질의를 하였기 때문에 실험은 총 18회 시도하였다. 시점 이동은 30회로 제한하였고 그 안에 찾지 못한 경우 실패로 정하였다. [그림 3]부터 [그림 8]은 실험의 예시를 나타낸다. [그림 3]은 입력 영상을 나타내며 흰 화살표는 Scan path를 나타낸다. Scan path는 입력 영상에서 구한 자극 영상의 밝기 값의 우선순위에 의해서 정해진다. [그림 4]는 예시 실험의 질의 영상으로 사용한 물체이다. [그림 5]는 기존 주목 알고리즘을 이용하여 계산한 자극 영상의 예시이다. [그림 6]은 [그림 5]의 자극영상을 바탕으로 Mask의 위치를 정해서 주목한 영상을 나타낸다. [그림 7]은 논문에서 제안한 질의 기반 주목 알고리즘을 적용해서 구한 자극 영상이며 자극도의 위치가 변한 것을 알 수 있다. [그림 8]은 [그림 7]의 결과에 Mask를 씌워 잘라낸 영상으로써 이 경우 시점 이동 1회에 검색에 성공하였다.

이번 실험을 통해 얻은 결과를 [그림 9]에 도시하였다. 또한 그래프를 분석한 자료를 [표 1]에 나타내었다.

4. 결론

질의 영상의 특징을 분석하여 주목 알고리즘에 적용하는 방법을 제시하고 실험을 통해 성능을 알아보았다.

[표 1]에서 질의 기반 주목 알고리즘이 기존 알고리즘보다 성능이 약간 향상되었음을 알 수 있다. 3회 이내의 시점이동으로 검색에 성공한 확률도 기존 방법에 비해 약 6% 증가하였다. 이 결과를 통해 질의 영상의 색상을 이용하여 주목 알고리즘의 자극도의 우선순위에 영향을 줄 수 있다는 것을 알 수 있었다.

그러나 [그림 7]을 살펴보면 12번째 실험에서 질의 기반 방법의 결과가 28번째로 가장 늦게 검색을 성공한 것으로 되어있다. 이것은 질의한 영상의 색상 분포가 R, G, B, Y의 영역에 들지 못하기 때문에 가중치를 정확히 부여하지 못하여 주목도의 우선순위가 잘못 변경된 것으로 추정된다. 특히 흰색이 많이 존재하는 질의 영상의 경우 대체적으로 늦게 검색하거나 검색에 실패하는 경향을 보였다([그림 7] 실험번호 14, 17). 색상 정보 외에 다른 속성의 정보를 질의 영상에서 추출하여 주목 알고리즘에 적용하면 이를 해결할 수 있을 것으로 보인다.

따라서 색상정보와 Shape정보를 동시에 사용하면 색상정보가 약한 경우에도 질의 물체에 대한 자극도 변화를 가능하게 할 수 있을 것으로 기대한다. 예를 들면 현재 본 논문의 질의 기반 주목 알고리즘은 Shape관련 정보로 외곽선과 Contour를 사용하였지만 질의에는 영향을 주지 않고 시점 이동시 같은 물체를 반복해서 검색하는 것을 방지하기 위해서만 사용하였다. 외곽선과 Contour에 대한 특징 영상을 새로 정의하고 이를 주목 알고리즘에 적용하면 질의 영상에서 Shape정보를 끌어내어 주목도 변화에 영향을 줄 수 있을 것으로 생각한다. 그러므로 추후에는 색상 정보 의외에 Shape와 관련한 성분을 분석하여 이를 주목 알고리즘에 적용하는 방법을 연구할 계획이다.

참고문헌

[1] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision 2004, January 5, 2004, pp.1-28

[2] Linda Lanyon and Susan Denham, "A Model of Object-Based Attention That Guides Active Visual Search to Behaviorally Relevant Locations,"

[3] J.J. Bonaiuto & L. Itti, "Combining attention and recognition for rapid scene analysis," Proceedings of the 2005 IEEE Computer Society Conference of Computer Vision and Pattern Recognition

[4] Dirk Walther, Ueli Rutishauser, Christof Koch and Pietro Perona, "On the usefulness of attention for object recognition," In Proc. WAPCV 2004

[5] Brad C. Motter, Eric J. Belky, "The guidance

of eye movements during active visual search," Vision Research 38, 1998, pp. 1805-1815

[6] Cheng Liu, "Gabor-Based Kernel PCA with Fractional Power Polynomial Models for Face Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 5, May 2004

[7] L. Itti, Christof Koch and Ernst Niebur, "A Model of Saliency-based Visual Attention for Rapid Scene Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, November 1998, Vol. 20, No. 11, pp.1254-1259