물체 인식을 위한 시각 주목 알고리즘

Visual Attention Algorithm for Object Recognition

류 광 근*, 이 상 훈**, 서 일 홍***
Gwang-geun Ryu, Sanghoon Lee, Il Hong Suh

Abstract - We propose an attention based object recognition system, to recognize object fast and robustly. For this we calculate visual stimulus degrees and make saliency maps. Through this map we find a strongly attentive part of image by stimulus degrees, where local features are extracted to recognize objects.

Key Words : Visual Attention, SIFT, Robot Vision, Object Recognition

1. 서 론

인간은 특정 물체를 찾고자 할 경우 뇌에 기억되어 있는 사전 정보를 검색하여 그 물체가 있을 만한 곳을 주로 찾아본다. 이 때, 인간은 자신이 움직이는 범위를 한정하기 위해 시야를 고정하게 되는데 이를 '주목(Attention)' 이라고 한다. 인간이 주목을 하기 위해서는 시각적인 자극과 함께 뇌의 기억들이 복합적으로 작용해야 한다. 인간의 집중이라는 뇌 활동을 로봇에 사용할 수 있도록 구현하여 로봇의 물체 인식 알고리즘에 적용한다[1].

로봇이 비전 센서로부터 얻을 수 있는 정보는 연산량에 따라 구분할 수 있다. 색이나 외곽선 정보는 적은 연산량을 갖지만 물체 인식을 위해 강인한 특징점을 구해야 할 경우 비교적 많은 연산을 필요로 하게 된다. 일반적으로 연산량은 특징점의 강인성에 비례하는데 이는 비주얼 정보의 특성상 픽셀 단위의 정보만으로 영상의 변화에 대해 모든 변환과정을 고려해야 하기 때문이다. 이러한 강인한 특징점 추출 알고리즘은 대표적으로 Scale Invarient Feature Transform (SIFT)[2]나 Harris Laplace Transform[3] 등이 있다. 특히 SIFT의 경우 각 특징점 위치에 대한 서술자(Descriptor)의 성능이 매우 뛰어난 것으로 알려져 있다[4]. 반면에 연산량이 많아 속도가 느린 것이 단점이다. 실제 800x400크기의 RAW 영상에 SIFT 연산을 수행할 경우 P4 2.8GHz시스템에서 특징점 추출에만 약 1.5초의 시간이 소요된다. 이렇게 얻어진 다량의 키포인트를 DB와 비교하기 위해서는 더 많은 시간이 소요된다. 따라서 이를 해결하기 위해서 비주얼 정보의 분할이 필요하다. 로봇의 비전센서에 들어오는 비주얼 정보를 인간이 주목을 하는 것과 마찬가지로 시각적 자극에 따라 범위를 한정하고 시점이동을 하게 하여 데이터의 양을 줄여서 특징점 추출에 걸리는 시간을 단축할 수 있다.

2장에서는 알고리즘을 구현하는 방법에 대해 언급하고 3장에서는 실험을 통해 알고리즘의 효용성을 분석한다.

저자 소개
* 류광근: 한양대학교 정보통신학과 석사과정
** 이상훈: 한양대학교 전자전기제어계측공학과 박사과정
*** 서일홍: 한양대학교 정보통신학과 교수

2. 시각 주목 알고리즘

인간의 주목은 기본적으로 흑백영상, 붉은색, 녹색, 방향성, 외곽선 등을 입력으로 받아 뇌에 저장된 사전 지식과 입력된 영상의 자극도 등을 고려하여 주목할 부분을 선택한다. 로봇의 주목 알고리즘도 기본적으로 이와 동일하다. 비전센서로부터 들어오는 정보는 컬러 영상으로부터 추출하는 색, 흑백영상, 방향성 등이며 사전 지식은 DB가 담당한다. 자극을 종합하여 자극영상(Saliency Map)을 만들면 주목할 위치를 선택할 수 있다. 주목을 할 위치를 주목 창(Attention Window (AW)라 한다. 주목 시스템 구성은 그림 1과 같다.

2.1 입력 영상 및 가우시안 피라미드 구성

카메라로부터 얻어지는 정보는 적색, 녹색, 파란색이다. 이 세 가지 정보로부터 밝기(I), 색깔(C), 방향성(O), 외곽선(E) 네 가지 정보를 수식(1)을 이용하여 얻을 수 있다.

$$\begin{aligned} I &= (r+g+b)/3 \\ R &= r-(g+b)/2 \\ G &= g-(r+b)/2 \\ B &= b-(r+g)/2 \\ Y &= (r+g)/2-|r-g|/2-b \\ O(0^\circ, 45^\circ, 90^\circ, 135^\circ) & \end{aligned} \quad -(1)$$

외곽선은 가우시안 차연산(Difference of Gaussian)을 통해 얻어내었다. 수식(1)의 결과를 이용하여 수식(2)와 같은 9개 층을 가지는 가우시안 피라미드를 구성한다.

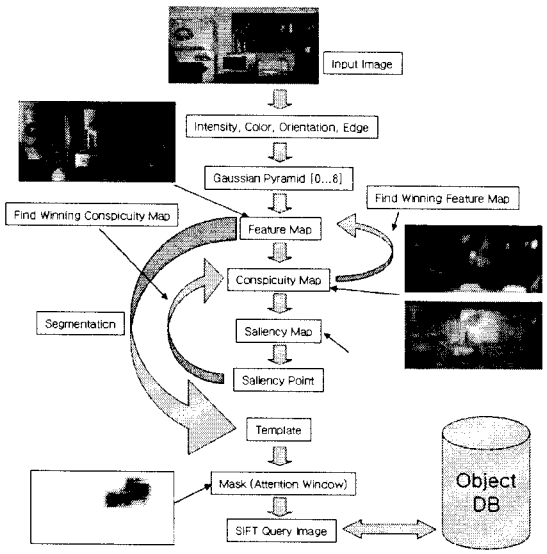

그림 1 전체 시스템 구성

$$I(\sigma), R(\sigma), G(\sigma), B(\sigma), Y(\sigma), O(\sigma,\theta), E(\sigma)$$
$$\sigma \in [0..8], \theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\} \quad -(2)$$

2.2 특징 영상 (Feature Maps)

가우시안 피라미드를 이용하여 특징 영상을 계산한다. 이때 중심주변차연산(Center Surround Difference(Θ))를 이용한다. 수식은 다음과 같다.

$$\begin{aligned} I(c,s) &= |I(c)\Theta I(s)| \\ RG(c,s) &= |(R(c)-G(c))\Theta(R(s)-G(s))| \\ BY(c,s) &= |(B(c)-Y(c))\Theta(B(s)-Y(s))| \\ O(c,s\theta) &= |O(c,\theta)\Theta O(s,\theta)| \\ E(c,s) &= |E(c)\Theta E(s)| \end{aligned} \quad -(3)$$
$$c \in \{2,3,4\}, s = c+\delta, \delta \in \{3,4\}$$

2.3 두드러짐 영상 (Conspicuity Maps)

Feature Map을 정규화(N)한 후 결합하여 Intensity($\bar{I}$), Color($\bar{C}$), Orientation($\bar{O}$), Edge($\bar{E}$), 4개의 Conspicuity Map 을 계산한다. 이를 위해 각 특징 영상의 층간 점들 간의 합을 구하는데 이를 층관통합(across-scale addition($\oplus$)) 이라 한다. 수식은 다음과 같다.

$$\begin{aligned} \bar{I} &= \bigoplus_{c=2s=c+3}^{4}\ \bigoplus^{c+4} N(I(c,s)) \\ \bar{C} &= \bigoplus_{c=2s=c+3}^{4}\ \bigoplus^{c+4} [N(RG(c,s))+N(BY(c,s))] \\ \bar{O} &= \sum_{\theta\in\{0^\circ,45^\circ,90^\circ,135^\circ\}} N\Big(\bigoplus_{c=2s=c+3}^{4}\ \bigoplus^{c+4} N(O(c,s,\theta))\Big) \\ \bar{E} &= \bigoplus_{c=2s=c+3}^{4}\ \bigoplus^{c+4} N(E(c,s)) \end{aligned} \quad -(4)$$

2.4 자극 영상 (Saliency Map)

위 4개의 두드러짐 영상을 수식(5)를 이용하여 자극 영상을 만든다.

$$S = \frac{1}{4}(N(\bar{I}) + N(\overline{C}) + N(\overline{O}) + N(\overline{E})) \quad -(5)$$

자극 영상은 영상의 자극도 분포를 표현한 것으로 값이 클수록 자극이 높다. 자극이 가장 높은 점을 최고 자극점(Saliency Point)라 하는데 이를 구하면 다시 두드러짐 영상으로 가서 색, 밝기, 방향성, 외곽선 두드러짐 영상들의 최고 자극점 좌표에 해당하는 점의 값을 구하여 가장 높은 값을 가지는 영상을 찾는다. 예를 들어 색 영상이 가장 높은 값을 가질 경우 색의 특징 영상 모임으로 올라가서 최고 자극점에 해당하는 점의 값 중 가장 높은 값을 가지는 특징 영상을 찾아낸다. 이 특징 영상은 주목창을 씌우기 위한 형틀의 기본이 된다. 특징 영상의 최고 자극점 주변으로 지역구분(Region Growing)과 적응 문턱(Adaptive Thresholding)을 통해 영상 구분(Segmentation)을 수행하고 입력 영상과 같은 크기로 확대한다. 최고 자극점에 해당하는 구역만 밝기값을 1로 하고 나머지를 0으로 하는 마스크(M)을 만든다. 이 마스크에 입력 영상을 Merge하여 SIFT 연산을 위한 질문 영상을 만드는데 수식(6)을 이용한다.

$$I'(x,y) = [255 - M(x,y)\cdot(255 - I(x,y))] \quad -(6)$$

질문 영상(I')를 자극점의 우선순위에 따라 계산하여 키포인트를 추출한 뒤 DB를 검색하여 물체 인식을 수행한다. 물체 인식을 수행하는 과정은 다음과 같다.

① 주목창 안에 있는 이미지에서 SIFT Feature를 구함.

② 주목창 안에 있는 키의 개수가 충분한지 검사 (5개 이상)

③ DB를 검색하여 각 Object 당 Match 된 수를 계산

④ 매치 된 물체 수가 많거나 매치 된 수가 부족할 경우 최고 자극점의 그 다음 우선순위로 위치를 옮겨서 AW의 크기를 늘림 (매치 됨 -> 5~7 개 이상 매치 된 경우)

⑤ 위의 과정 반복. 물체 인식할 경우 종료

3. 실험 및 결과

실험의 목적은 집중을 하지 않은 전체 이미지를 넣어서 SIFT 특징점을 추출하여 DB를 검색, 물체 인식을 수행하는 일련의 과정을 집중을 적용한 이미지를 작성하여 같은 과정을 수행하게 한 뒤 물체 인식의 정확도와 속도를 비교한다. 실험은 펜티엄4 2.8GHz, 1GB 메모리를 장착한 컴퓨터에서 진행하였으며, 실험에 사용한 영상은 800x400x24bit의 RGB 영상이다. DB에 저장한 물체의 개수는 총 11개이다. 실험은 2가지 카테고리로 구분하여 진행한다.

3.1 물체 인식 속도

전체 이미지를 넣어서 키포인트를 추출한 뒤 DB를 검색하는 시간과 집중을 통해 같은 과정을 수행하여 물체 인식을 하도록 한다. 집중을 적용할 경우에는 AW의 사이즈를 늘려가며 키포인트의 키 개수를 충분히 확보하는 과정도 포함된다. 실험 시도 횟수는 총 10회이다. 결과는 표1과 같다. 집중

을 적용하였을 경우 DB검색시간이 늘어나는데 그 이유는 AW크기를 조절하여 충분한 양의 특징점을 확보하기 위해 반복하기 때문이다. 최소 5회의 반복을 필요로 하였으며 평균적으로 10회 전후에 물체 인식을 하였다.

	특징점 추출 평균 시간
집중 비적용	1352ms
집중 적용	836ms

표1. 물체 인식 속도 결과

3.2 물체 인식 정확도

물체 인식은 두 가지 경우로 나눌 수 있다. 첫 번째는 질문 영상에서 추출한 특징점이 DB에 저장된 어떤 물체의 특징점과 5개 이상 매치 되었을 경우 매치에 성공하였다고 한다. 두 번째로 매치에 성공한 물체가 질문 영상의 그것과 같을 경우 이를 올바로 매치되었다고 한다. 실험 결과를 표 2에 나타내었다.

정답 비율은 실험을 한 총 횟수에 대해 올바른 매치 결과를 나타낸 횟수를 DB안에 있는 전체 물체의 개수로 나눈 것이다. 총 실험은 10회 수행하였으며 DB에 저장된 물체는 11개이다.

	매치 성공	올바른 매치
집중 비적용	0.325	0.25
집중 적용	0.4	0.6

표2. 물체 인식 정확도 결과

4. 결론 및 고찰

로봇의 물체 인식을 위한 집중 알고리즘의 효용성을 실험을 통해 검증하였다. 실험 결과 집중 알고리즘을 적용하면 물체 인식의 정확도가 증가하는 것을 알 수 있었다. 집중 알고리즘의 장점은 무엇보다 특징점의 강인성을 포기하지 않고 특징점의 개수만을 줄여 검색시간을 최소화 하는 것과 물체 영역 제한을 통해 인식 정확도를 높일 수 있다는 것이다. 실험 1에서 AW생성까지 걸리는 시간을 줄이기 위하여 전처리 과정을 고속화하거나 혹은 다른 프로세스를 통해 미리 실행시키는 작업이 필요하다.

현재 집중 알고리즘은 점 단위의 자극도에 따라 AW를 선택한다. 비록 자극도를 색과 밝기, 방향성 등 여러 성격에 따라 분류를 하였지만 물체 단위의 특성을 표현했다고 보기 힘들다. 따라서 물체를 구분할 수 있는 새로운 특징 영상을 도입하여 물체 기반으로 시점 이동을 할 수 있게 해야 할 것이다. 그리고 실제로 로봇에 적용해야 하기 때문에 로봇의 이동, 카메라 회전 등의 로봇의 움직임과 연계할 수 있는 물체 인식 집중 알고리즘 시스템을 구현할 계획이다.

참 고 문 헌

[1] Linda Lanyon and Susan Denham, "A Model of Object-Based Attention That Guides Active Visual Search to Behaviourally Relevant Locations," WAPCV2004, LNCS 3368, pp.42-56, 2005

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

[3] Krystian Mikolajczyk and Cordelia Schmid, "Indexing based on scale invariant interest points," Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference, Vol. 1, pp. 525-531

[4] Krystian Mikolajczyk and Cordelia Schmid, "A Performance Evaluation of Local Descriptors," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 10, October 2005, pp.1615-1630

[5] J.J. Bonaiuto and L. Itti, "Combining Attention and Recognition for Rapid Scene Analysis," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005

[6] Dirk Walther, Ueli Rutishauser, Christof Koch and Pietro Perona, "On The Usefulness of Attention for Object Recognition," 2nd Workshop on Attention and Performance in Computational Vision at the European Conference for Computer Vision, 2004, pp.96-103