

적응적인 Saliency Map 모델 구현

박상범#, 김기중*, 한영준*, 한헌수*

Implementation of Image Adaptive Map

Sang Bum Park#, Ki Joong Kim*, Young Joon Han* and Hern Soo Hahn*

ABSTRACT

This paper presents a new saliency map which is constructed by providing dynamic weights on individual features in an input image to search ROI(Region Of Interest) or FOA(Focus Of Attention). To construct a saliency map on there is no a priori information, three feature-maps are constructed first which emphasize orientation, color, and intensity of individual pixels, respectively. From feature-maps, conspicuity maps are generated by using the It's algorithm and their information quantities are measured in terms of entropy. Final saliency map is constructed by summing the conspicuity maps weighted with their individual entropies. The prominency of the proposed algorithm has been proved by showing that the ROIs detected by the proposed algorithm in ten different images are similar with those selected by one-hundred person's naked eyes.

Key Words : Saliency map (현저한 지도), Feature-map (특징 지도), Conspicuity map (현격함 지도), Region of interest (관심 영역), Focus of attention (관심 집중)

1. 서론

인간은 자신이 관심을 가진 소리나 보고 싶은 대상에 특별한 관심을 가지고 귀 기울이거나 시선을 집중하게 된다. 즉, 작은 소리라도 자신이 원하는 사람의 소리는 집중하여 그 소리를 들으려고 애를 쓰게 되고, 눈에 보이는 장면에서 자신이 찾고자 하는 특징을 집중하여 보게 된다. 이와 같이 매체가 제공하는 정보에서 특정하게 선택적으로 주의를 집중하여 인지하려는 심리현상을 주의집중(attention)이라 한다.¹ 주의집중이 이루어지면 일반적으로 집중하려는 대상을 제외한 다른 정보는 자

연스럽게 무시하거나 인지하려는 대상에서 제외하게 된다. 예를 들면 방에서 대화할 때 다른 사람의 대화는 무시하고 상대방의 말을 주의 깊게 들을 수 있는 것과 같은 것이다. 주의집중은 운전하면서, 화장을 하고, 동시에 휴대 전화로 통화하는 것과 같이 여러 개로 분리 될 수도 있다.

인간의 눈은 동시에 입력되는 많은 시각 정보를 두뇌로 전달하고, 인간의 신호 전달 체계가 가지는 전달 속도의 한계에도 불구하고 복잡한 시각 정보를 실시간으로 처리하는 탁월한 능력을 가지고 있다.² 인간의 시각 시스템은 선택적 주의집중에 의해서 시각 수용체로 도달되는 많은 물체들

☞ 접수일: 2007년 10월 5일; 게재승인일: 2007년 12월 28일

교신저자: 숭실대학교 전자공학과

E-mail: forcepsb@ssu.ac.kr Tel. (02) 821-2050

* 숭실대학교 전자공학과

중에서 필요한 정보만을 추출하여 원하는 작업을 수행하며, 각 기능을 수행하는 결과를 얻을 수 있다. 영장류에 대한 실험적 증거에 의하면 입력되는 영상에서 가장 관심이 있는 부분을 찾아내는 “시각 정보 선택(visual selection)” 기능에 관련된 수많은 메커니즘이 존재한다고 알려져 있으며, 시각적 주의 기능이 이에 해당하는 것이라 할 수 있다.³

시각적 주의를 제어할 수 있는, 신경계를 모방한 최초의 계산적 모델은 1985년 Koch 등⁵에 의해 만들어졌다. 이들이 제시한 모형을 바탕으로 많은 모델이 제안되었으며 이 중에서 주목할 만한 모델은 Milanese⁶와 Itti 등²이 제안한 현저한 지도(Saliency map) 모델이라 할 수 있다. 이러한 모델들은 현재 시각적 주의에 대한 개념 정립 단계에 있으며, 특정 문제에 한해서만 동작하거나 실질적인 측면이 아닌 이론적인 측면에 많이 치중되어 있기 때문에 실제적인 문제에 제대로 적용되지 못하고 있는 등의 문제를 안고 있다.³ 이러한 문제를 해결하기 위해 다양한 연구가 시도되고 있으며, 특히 Itti 등이 제안한 모형인 현저한 지도를 개선시키기 위한 많은 연구들이 진행되었다. 손준일 등은 독립성분 해석 알고리즘(Independent-Component Analysis)을 적용시킨 현저한 지도 모델을 제시하였고,⁷ Nabil 등⁸은 비선형방식으로 현저한 지도를 만들었다. 또한, Nabil 등^{9,10}은 각각의 현격함 지도(Conspicuity map)에 가중치를 적용한 현저한 지도 모델을 제시하기도 하였다. 이러한 현저한 지도 모델은 시간 특징 중 하나인 모션정보를 단서로 사용하여 동영상에서도 사용되어질 수 있도록 확장되어진다. 박민철 등¹¹은 모션정보만을 이용하여 현저한 지도를 구성하고 이것을 인간을 대상으로 한 심리학적 실험과 비교를 하였다. 또한, Nabil 등¹²은 동적 현저한 지도란 모델을 제시하고 이것을 정적 현저한 지도와 조합시켜 동영상에서도 효율적인 새로운 현저한 지도 모델을 제시하였다. 그러나 위에서 제시한 모델은 색상이 함유된 영상에서만 실험을 하였고 임의로 특정 현격함 지도에 가중치를 부여하거나, 현격함 지도의 단순 합으로 현저한 영역을 검출하였다.

이와 같이 기존에 제한된 방법들이 가지는 문제들을 고려하여 본 논문에서는 ROI(region of interest) 혹은 FOA(focus of attention)를 탐지하기 위해 능동적 가중치를 부여하는 적응적인 현저한 지도 모델을 제시하고 이를 검증한다. 영상에서 현저

한 지도를 만들기 위하여 방향정보, 색상정보, 그리고 밝기 정보를 기반으로 하여 현격함지도를 만든다. 현격함 지도는 일반적으로 사용하는 Itti가 제안한 기법²을 이용하여 만들고, 각각의 현격함 지도의 특징을 부각시키기 위하여 동적 가중치(dynamic weight)를 부여한 현저한 지도를 만들어낸다. 동적 가중치는 현격함 지도에 Chang¹⁴이 제안한 엔트로피 기법을 적용시켜 영상의 정보량을 측정 한 후, 이에 따라 계산된 가중치를 각각의 현격함 지도에 부여한다. 이렇게 가중치를 부여받은 현격함 지도를 선형 조합함으로써 만들어진 현저한 지도는 설문조사를 통한 비교로 그 성능을 검증한다.

본 논문은 다음과 같이 구성된다. 2 장에서는 본 논문에서 제안하는 현저한 지도 모델을 구성하는 방법을 설명하고, 3 장에서는 제안한 모델의 우수성을 검증하기 위한 실험방법 및 결과를 설명한다. 끝으로 4 장에서 결론 및 향후 연구 방향에 대해 설명한다.

2. 적응적인 현저한 지도 모델

현저한 지도를 구성하는 방법은 영상을 관찰하는 관찰자가 관심을 가진 영역을 찾는 하향식 접근방법과 영상에 대한 사전지식 없이 주어지는 시각 자극의 특징을 기반으로 관심영역을 추출하는 상향식 접근방법이 있다. 본 논문에서는 임의의 영상이 주어지는 경우, 관찰자가 사전에 의도하는 바가 없이 영상의 시각적인 특징만을 참고하여 관심영역을 결정하는 상향식 접근방법을 제안한다. 본 논문에서 새롭게 제시하는 모델은 Fig. 1에서 제시한 과정을 거쳐 현저한 지도를 만든다. 입력되는 영상은 시각적 주의를 집중시키는 특징들로 구성된 여러 특징지도로 분리된다. 본 논문에서는 영상의 색상과 명암도, 그리고 움직임 정보를 특징으로 사용하였다. 이 세 가지 특징을 표현할 수 있는 특징지도로 다중 피라미드 영상을 생성한 후, 생성된 다중 피라미드 영상의 상호간의 차이(center surround difference)를 구한다. 얻어진 차이정보는 각 기본 요소가 주변으로부터 얼마나 두드러졌는지를 나타내게 된다.

차이정보를 이용하여 현격함 지도를 만들고 현격함 지도에 PIM(Picture Information Measure)¹³을 이용하여 각각의 정보량을 추출한 후, 정보량에

따라서 동적 가중치를 부여한다. 현저한 지도는 부여받은 동적 가중치에 따라서 선형적으로 더해져 만들어진다.

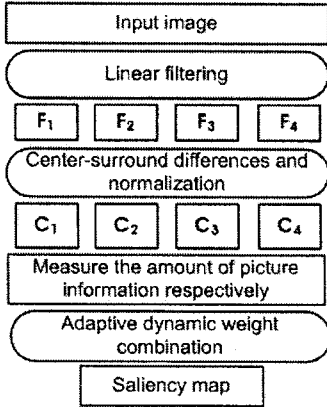


Fig. 1 Flow of saliency map construction

2.1 특징지도

인간의 망막은 감광성 요소가 공간적으로 비균일하게 배치되어 있어 시각정보를 받아들일 때 다른 영역과 구별되는 특정영역이나 물체에 시선을 집중하도록 하여 그 부분에 대하여 더욱 자세한 분석이 가능하도록 되어있다. 정신 물리학적 실험에 의하면 인간은 주의가 가는 영역을 여러 개의 상향식 기준에 의해 재빠르게 탐지할 수 있다고 한다.³ 이러한 특성을 고려하여 본 논문에서는 사람이 색상과 명암도에 따라 주의가 집중된다고 판단하고 색상과 명암도의 특징지도를 사용하였다.

2.1.1 색상 특징지도

인간의 색상 지각 능력은 망막에 있는 추상체 (cone)들의 반응에 의해 가능하다. 인간의 눈에는 여러 가지 파장의 빛을 3 개의 색상으로 지각할 수 있는 3 가지 종류의 추상체가 있으며, 이것은 적색, 녹색, 청색의 빛 중 어느 하나에 반응하는 광색소를 가지고 있다. 이러한 추상체의 반응은 신경절 세포에 넘겨지고 바로 LGN(lateral geniculate nucleus)을 거쳐 색상정보를 뇌로 전달한다. 색상정보를 뇌로 전달해 주는 신경회로는 3 가지 추상체들에 의한 정보를 ‘적/녹’, ‘황/청’의 반대 쌍의 색상정보로 바꾸어 전달한다.¹³

이러한 생물학적인 시각처리 기관의 특성을 고려하여 본 논문에서는 다음과 같은 순서에 따라

색상특징지도를 만든다. 먼저 r, g, b 로 표현되는 적색, 녹색, 청색을 식 (1)을 사용하여 널리 조율된 4 개의 색상채널(R 은 적색채널, G 는 녹색채널, B 는 청색채널, Y 는 황색채널)로 변환한다.²

$$\begin{aligned} R &= r - (g + b) / 2, \\ G &= g - (r + b) / 2 \\ B &= b - (r + g) / 2, \\ Y &= r + g - 2(|r - g| + 2) \end{aligned} \quad (1)$$

각 채널은 흑색과 백색에 대해 0 을 출력한다. 이 채널을 사용하여 두 개의 색상을 특징으로 하는 특징지도를 식 (2)와 같이 구성한다.

$$\begin{aligned} RG_{x,y} &= R_{x,y} - G_{x,y} \\ BY_{x,y} &= B_{x,y} - Y_{x,y} \end{aligned} \quad (2)$$

여기서 x, y 는 영상의 x 축, y 축 좌표를 나타낸다.

Fig. 2(b)와 2(c)는 Fig. 2(a)에 주어진 입력영상에서 식 (2)를 이용하여 추출한 색상 특징지도를 보여준다. RG feature 는 주로 적색을 추출하고 BY feature 는 주로 청색을 추출하게 된다. 따라서 Fig. 2(b)는 적색이 두드러지는 깃발을 feature 로 추출하였고, 2(c)는 청색이 그나마 있는 가이드 부분에서 feature 가 추출되었다.

2.1.2 명암도 특징지도

명암도 영상은 실제적으로 현저한 특징을 추출하는데 중요한 역할을 하므로 식 (3)을 이용하여 명암도 영상을 만들고 이 영상을 특징지도로 사용한다.

$$I = (r + g + b) / 3 \quad (3)$$

Fig. 2(d)는 Fig. 2(a)에 주어진 입력영상에서 식 (3)을 이용하여 추출한 명암도 특징지도를 보여준다.

2.1.3 방향 특징지도

영상에서 화소의 방향특징을 나타내는 $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ 의 4 가지 방향성정보를 Gabor filter 를 사용하여 획득한다. Gabor filter 는 2 차원 가우시안과 코사인 격자의 합으로 구성되어지는 필터로 영상으로부터 주파수 성분과 방향성을 가지는 영상 구조가 어디에 있는지의 지역적인 정보를 효과적

으로 표현할 수 있다. Fig. 2(e)-(h)는 Fig. 2(a)에 주어진 입력영상에 대하여 구한 4 방향 특징지도를 보여준다. Fig. 2(e)는 0°의 특징을 추출하였고, Fig. 2(f)는 45°의 특징을 Fig. 2(g)는 90°의 특징을, Fig. 2(h)는 135°의 특징을 추출하였다. Fig. 2는 우리의 시각 시스템과 유사하게 영상에서 90° 성분이 가장 적게 나타난 결과를 확인할 수 있다. 만약 입력 영상이 빌딩이 많은 영상이라면 90° 도에 해당하는 성분이 많이 나타나게 될 것이다. 이렇게 획득된 4개의 특징지도는 하나의 현격함 지도로 합쳐진다.

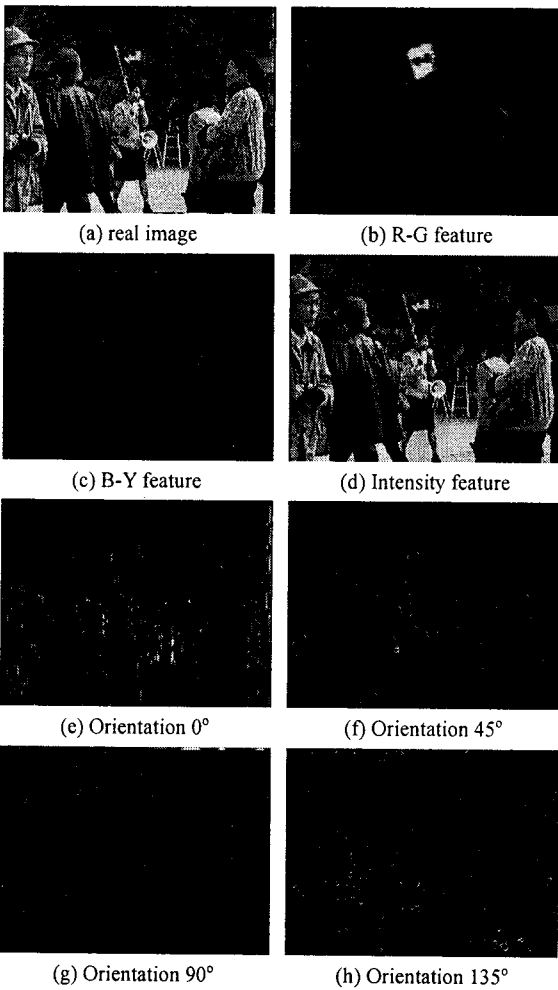


Fig. 2 Example of feature map construction

2.2 현격함 지도

주어진 특징지도는 사용한 특징이 영상에 어떻게 분포되었는지를 보여준다. 이로부터 영상의 어

느 영역이 주변과 대비하여 주어진 특징을 두드러지게 보여주는 지를 강조하는 현격함 지도를 만들게 된다. 생물학적으로 가장 타당한 모델은 중심-주변 메커니즘이다.² 이 방법은 영상의 다중 해상도 표현에 기반을 두고 있다. 각 해상도 특징을 j 라 하면, 9 개의 크기 가우시안 피라미드 P_j 는 j 가 증가함에 따라 2의 배수로 subsampling 된 F_j 에 lowpass 필터중에 가장 널리 쓰이는 Gaussian filter를 적용하여 획득된다. 식 (4)는 특징지도 F_j 에 가우시안 필터 G 를 사용하여 점진적으로 subsampling 하고 lowpass 필터링하는 과정을 보이고 있다.

$$P_j(0) = F_j$$

$$P_j(i) = F_j(i-1) * G \quad (4)$$

여기서 *는 공간 컨볼루션 연산자이며, j 는 해상도의 인덱스를 나타내고 i 는 각 특징지도를 나타낸다.

가우시안 피라미드를 취하여 각 영상을 확대하면 가장 주요한 부분만 남게 된다. 이러한 특징을 이용하여 중간 다중 스케일 현격함 지도를 만든다. 각각의 특징에 대해서, 6개의 다중스케일 현격함 지도들인 $M_{j,k}(1...6)$ 를 식 (5)을 이용하여 구한다.

$$M_{j,1} = |P_j(2) - P_j(5)|,$$

$$M_{j,2} = |P_j(2) - P_j(6)|,$$

$$M_{j,3} = |P_j(3) - P_j(6)|,$$

$$M_{j,4} = |P_j(3) - P_j(7)|,$$

$$M_{j,5} = |P_j(4) - P_j(7)|,$$

$$M_{j,6} = |P_j(4) - P_j(8)| \quad (5)$$

이 중간 다중스케일 현격함 지도들은 공간 주파수들의 차이에 민감하다. $M_{j,1}$ 과 같은 fine map은 고주파 영역을, $M_{j,6}$ 같은 coarse map은 저주파 영역을 각각 찾아낸다. 각각의 특징 j 에 대해서 6개의 중간 다중 크기 현격함 지도들은 유일한 특징을 가지는 현격함 지도 C_j 로 식 (6)에 의해 합쳐진다.

$$C_j = \sum_{k=1}^6 N(M_{j,k}) \quad (6)$$

현격함 지도로 합쳐질 때 6 개의 중간 다중 크기 현격함 지도들은 서로 다른 범위의 값을 가지고 있기 때문에 동등한 비교를 위해 정규화를 수행하였다. $N(\bullet)$ 는 정규화 함수를 보여주고 있다. 정규화는 다음 식 (7)과 같이 정의된다.

$$F_{(x,y)}^{new} = \frac{F_{(x,y)} - \min}{\max} \quad (7)$$

위 식에서 min 과 max 는 각각 영상에서 가장 큰 명암도값과 가장 작은 명암도값이다. Fig. 3 은 Fig. 2 에서 주어진 각 특징지도들을 현격함 지도로 변환한 결과 영상을 보여준다.

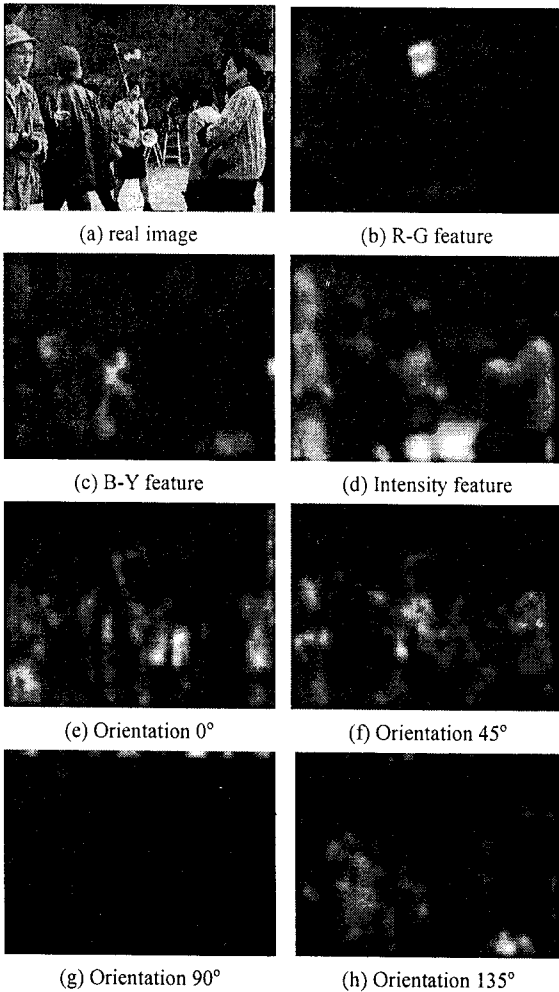


Fig. 3 Conspicuity maps extracted from the feature maps given in Fig. 2

2.3 현저한 지도

시각적 주의모델인 현저한 지도는 시각적인 환경에 놓여있는 특이할만한 물체들에 대한 정보를 가지고 있는 2 차원지도로 의미하며, 입력영상의 여러 특징들을 강조하도록 구성된 현격함 지도를 융합하여 생성한다. 현저한 지도 X 는 식 (8)과 같이 4 개의 특징(R-G feature, B-Y feature, Intensity feature, Orientation feature)에서 얻어진 현격함 지도들에 특정한 가중치 α_i 를 부여하고 합하여 만들어진다.

$$X = \sum_{i=1}^4 \alpha_i C_i \quad (8)$$

α_i 는 각 현격함 지도들의 특성에 따라 동적으로 결정되기 때문에 동적가중치라 하며 이 함수에 따라 현저한 지도의 성능이 결정되는 중요한 의미를 갖는다. 따라서 이를 결정하는 방법은 다음 절에서 별도로 설명하기로 한다.

2.4 동적 가중치

현격함 지도는 이를 생성하는 특징값을 두드러지게 표현한다. 따라서 특징별로 얻어진 현격함 지도가 강조하는 영역은 각각의 현격함 지도마다 매우 독립적인 성격을 지니고 있기 때문에, 각각 다른 영역을 현저한 영역이라고 강조하게 된다. 이러한 이유로 영상의 각 지점마다 하나의 중요도 측정치를 나타내주는 현저한 지도를 만들기 위해서 이들 현격함 지도들을 최적의 비율로 결합하여야 한다.

일반적으로 현저한 지도는 모든 현격함 지도에서 얻어지는 전체 특징값 중 가장 큰 값을 취하거나, 각 특정한 임계치보다 큰 현격함 지도를 결합하는 방법으로 구성하였다. 혹은 각각의 현격함 지도에 적절한 가중치를 두어 합하는 방법이 사용되기도 하였다. 결국, 가장 흥미로우면서도 중요한 곳을 선택하기 위해서는 가장 현저하다는 것을 나타내는 전역적인 하나의 현저함을 나타내는 중요도 측정치를 얻어야 하며, 이러한 측정치에 따라 특징들을 결합하여 하나의 유일한 현저한 지도를 만들어야 한다.³ 이를 위해 본 논문에서는 영상의 엔트로피를 측정하여 동적 가중치를 할당하는 방법을 제안한다.

본 논문에서는 엔트로피를 정량화하기 위해

Shannon의 엔트로피 개념을 이미지 정보에 적용시킬 수 있도록 Chang¹³이 제안한 PIM(Picture Information Measure)을 사용한다. 식 (9)와 같이 정의되는 PIM은 해당 블록의 전체 화소 수와 히스토그램 중 가장 큰 도수의 히스토그램 값과의 차이로 계산된다.

$$PIM = \sum_{i=0}^{L-1} h(i) - \text{Max}_i(h(i)) \quad (9)$$

여기서 L은 양자화 레벨 수, h(i)는 각 해당 블록(또는 영상)에서 레벨 i에 대한 히스토그램을 의미하며, Max_i는 각 해당 블록(또는 영상)의 가장 큰 도수의 히스토그램값을 나타낸다.

PIM은 해당 블록이 많은 정보를 갖는 경우에는 큰 값을 갖게 되고, 적은 정보량을 갖는 경우에는 작은 값을 갖게 된다. Fig. 4는 영상의 특징에 따른 20x20 블록으로 측정된 PIM 값을 보여준다. Fig. 4(a)의 평균과 분산값은 각각 M=331.7448, V=1274.9536으로 획득되었으며, Fig. 4(b)는 M=243.9167, V=9980.9854으로 획득되었다.

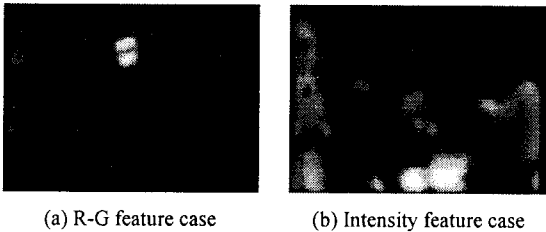


Fig. 4 Average(M) and variance(V) depending on the used features

전체적으로 복잡한 블록들이 비교적 균일하게 분포하는 Fig. 4(a)에서 PIM 값들의 평균은 크고, PIM 값들의 분산은 작은 것을 볼 수가 있다. 영상에서 PIM의 평균이 크다는 것은 영상에 엔트로피가 큰 영역이 많다는 것을 의미하고, PIM의 분산이 작다는 것은 영상에 전체적으로 균일한 크기의 엔트로피 성분이 분포하고 있음을 말해준다. 이와는 대조적으로 Fig. 4(b)를 보면 Fig. 4(a)에 비해 PIM 값의 분산이 상대적으로 크고 평균은 작은 것을 볼 수 있다. 분산이 큰 이유는 영상 내에 엔트로피가 높은 성분과 낮은 성분이 적당량 이상 상존하고 있기 때문이다. 즉, 엔트로피가 비교적 넓은 분포의 히스토그램을 갖기 때문에 큰 분산 값을 갖게 된다.

또한 비교적 단순한 블록들 위주로 구성되어 있기 때문에 PIM 값들의 평균은 작은 값을 갖는다. 이 PIM 값은 계산의 편리성을 위해 그 값을 0~1 값으로 정규화한다.

앞에서 언급한 바와 같이 영상에서 고정된 크기의 하위 블록들로부터 NPIM(Normalize Picture Information Measure)을 구하여 그 값들의 평균과 분산을 취하여 영상을 해석하는데, 본 논문에서는 영상을 20x20 크기의 블록으로 나누어 각각의 블록에 대한 NPIM을 계산한 후 영상 전체에 대한 NPIM의 평균과 분산을 구하였다. 이 두 값은 현저한 지도 구성요소를 선택하기 위한 기준으로 사용된다. NPIM은 식 (10)과 같이 정의된다.

$$NPIM = 1 - \frac{\text{Max}_i(h(i))}{\sum_{i=0}^{L-1} h(i)} \quad (10)$$

이를 이용하여 동적 가중치를 구하는데 있어서 본 논문에서는 3가지 가정을 전제하고 있다.

1. 현격함 지도는 가우시안 밀도를 가지는 랜덤변수이다.
2. 현격함 지도는 서로 독립적이다.
3. 현저한 지도의 평균과 분산을 알고 있다.

가정 1은 현저한 지도를 식 (8)과 같이 랜덤변수로 표현되는 현격함 지도 N개의 가중합(weighted sum)으로 간주하는 것이다. α_i는 실수의 가중상수(weighted constant)이다. 이를 이용하면 X의 분산은 다음과 같이 구할 수 있다.

$$\begin{aligned} E[X] &= \sum_{i=1}^N \alpha_i E[X_i] = \sum_{i=1}^N \alpha_i \bar{X}_i \\ X - \bar{X} &= \sum_{i=1}^N \alpha_i (X_i - \bar{X}_i) \\ \sigma_x^2 &= E[(X_i - \bar{X}_i)^2] \\ &= E\left[\sum_{i=1}^N \alpha_i (X_i - \bar{X}_i) \sum_{j=1}^N \alpha_j (X_j - \bar{X}_j) \right] \\ &= \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j E[(X_i - \bar{X}_i)(X_j - \bar{X}_j)] \\ &= \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j C_{X_i X_j} \end{aligned} \quad (11)$$

따라서, N개의 랜덤변수 X_i(가중값 α_i)의 가

중합의 분산은 그들의 공분산 C_{X_i, X_j} (가중값 α_i, α_j)의 가중합과 같다. 여기에서 가정 2를 고려하면 공분산 C_{X_i, X_j} 과 현저한 지도의 분산은 각각 식 (12)와 식 (13)로 표현된다.

$$C_{X_i, X_j} = \begin{cases} 0 & i \neq j \\ \sigma_{X_i}^2 & i = j \end{cases} \quad (12)$$

$$\sigma_X^2 = \sum_{i=1}^N \alpha_i \sigma_{X_i}^2 \quad (13)$$

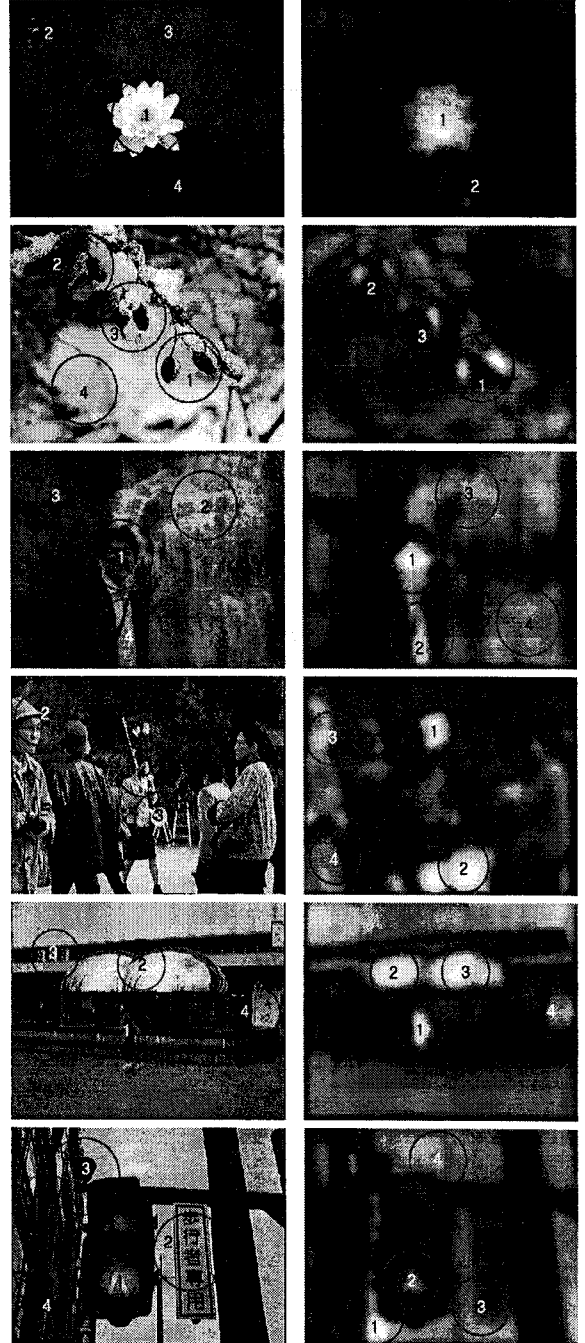
위의 정의를 이용하여 랜덤변수의 평균과 분산의 가중합이 현저한 지도의 평균과 분산이라 가정한다. 우리가 알고 있는 식은 분산과 평균에 관한 식 두 개이나 구해야 할 변수는 4 개이다. 따라서 본 연구에서는 식 (10)와 식 (13)를 기반으로 하는 각각의 현격함 지도에서 나타난 결과를 더하여 하나의 map 을 생성한 뒤 각 결합쌍의 가중치를 구하였다. 각 쌍의 가중치는 다시 현격함 지도의 가중치를 부여하기 위한 값으로 쓰인다. 현저한 지도의 평균과 분산은 경험치를 적용하였다.

3. 실험

제안하는 방법을 통하여 구성된 현저한 지도의 성능은 제안하는 알고리즘이 선택한 영역과 다수의 사람들이 선택한 영역과 일치하는 지를 비교하여 평가하였다. 성능평가 자료를 만들기 위하여 10 개의 영상을 선정한 후, 100 명에게 각각의 영상을 보여주며 가장 두드러지는 곳의 우선순위를 설문조사하여 데이터를 만들었다. 그리고 제안하는 방법을 통하여 현저한 지도를 만들어 그 결과를 비교하였다. 제안하는 방법은 VC++로 구현하였다.

Fig. 5 의 왼편은 실험에 사용한 10 개의 영상에 100 명이 선택한 관심영역을 우선순위로 숫자로 표시한 영상을, 오른쪽엔 제안한 알고리즘을 이용하여 구성된 현저한 지도를 보여준다. 현저한 지도에도 값의 크기에 따라 큰 값을 갖는 영역부터 숫자로 표시하여 사람이 선택한 영역과 비교할 수 있도록 하였다. 입력영상이 인물영상이거나 특정한 물체를 대상으로 획득한 경우(Fig. 5 에서 1, 3, 4, 8 번째 영상), 사람이 선택한 1 위영역의 경우 대부분 80 명 이상이 선택하였다. 그러나 일반적인 대중이나 자연영상의 경우(Fig. 5 에서 2, 4, 7 번째 영

상), 관측자의 성별이나 나이 등의 특성에 따라 1 위에서 4 위를 고르게 선택하는 경향을 보였다.



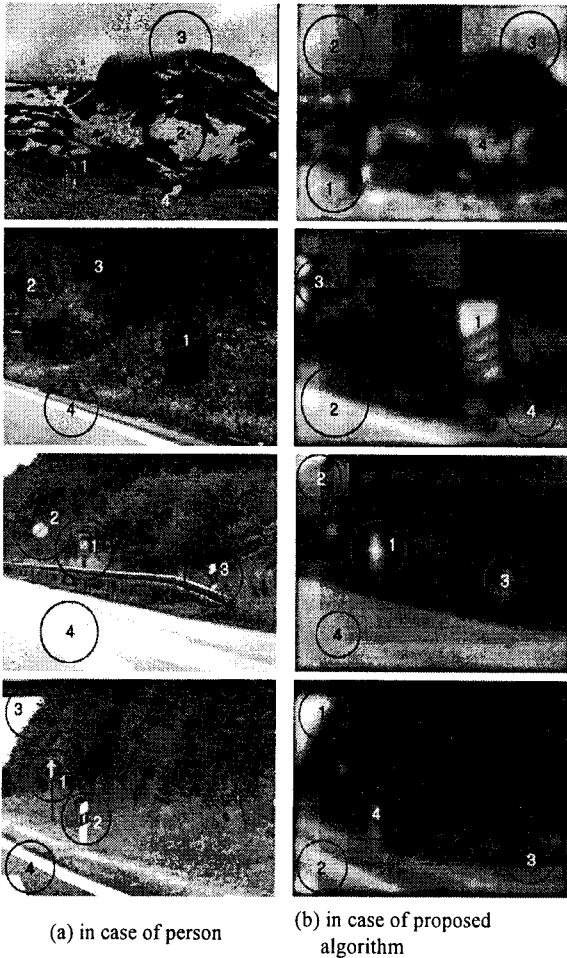


Fig. 5 Comparison of Regions of Interest selected by 100 people and the proposed algorithm

Table 1 Detection ratios between ROIs which are selected by both 100 people and the proposed algorithm

Algorithm	1st	2nd	3rd	4th	avg
Human					
1st	8	1	3	1	-
2nd		1	3	5	-
3rd	1				-
4th	1	4		2	-
Selected rate(%)	100	60	60	80	75

Table 1 은 제안한 알고리즘이 선택한 관심영역과 100 명의 사람이 동일한 영상에서 선택한 관심

영역과의 일치 여부를 표로 작성하여 보여주고 있다.

Fig. 5 의 실험결과에서 알 수 있듯이 대부분의 영상에서 알고리즘이 선택한 1 번 영역은 사람과 일치하는 것을 알 수 있다. Table 1 에서 보면 알고리즘이 첫 번째 관심을 가진 영역은 사람도 대부분 관심영역으로 선정함을 확인할 수 있다. 알고리즘이 그 이하의 우선 순위로 선택한 관심영역도 대부분 사람이 관심영역으로 선택하였으나 6 번과 9 번째 영상에서와 같이 영상의 특성에 따라 전혀 선택되지 못하는 경우도 발생하였다. 이 예에서 보듯 사람들이 현저한 영역이라 생각하는 부분이 다양하기에 본 논문에서 제시한 현저한 지도가 완벽하게 모든 현저한 영역을 검출한다고 볼 수는 없지만 첫 번째로 선택한 관심영역의 경우 100% 일치하였고 전체적으로 75% 이상의 정합되는 결과를 고려할 때 대체적으로는 우수한 성능을 보이고 있음을 확인할 수 있었다.

4. 결론

본 논문에서는 미지의 영상이 주어지는 경우 인간은 영상에서의 연속성, 크기, 배경과의 차별성에 시선을 집중하고 있음에 착안하여, 현저한 지도를 만들기 위해, 각 화소의 방향, 색상 및 밝기 정보별로 특징지도를 만들고 이들 영역에서 Itti 의 알고리즘을 이용하여 현저함 지도를 생성한 후, 각 map 에 포함된 entropy 를 가중치로 활용하여 결합하는 방법을 제안하였다. 가중치는 각 영상의 엔트로피를 PIM 으로 계산하여 동적으로 할당함으로써 영상의 특징을 잘 살릴 수 있음을 확인하였다. 제안된 알고리즘의 성능은 10 장의 서로 다른 특징을 갖는 영상에 대해, 실제 사람들이 선택한 관심영역과 알고리즘이 추정한 관심영역과의 일치 여부를 가지고 판단하였으며, 전체적으로 75% 이상의 일치율을 보임으로써 제안한 알고리즘이 우수함을 확인할 수 있었다. 이와 같은 연구결과는 향후 자동감시 시스템이나 미지의 공간을 탐사하는 경우 의미있는 영역을 선택하는 작업에 활용하기 위해 활용할 예정이다.

후 기

본 연구는 숭실대학교 교내연구비 지원으로

이루어졌음.

참고문헌

1. Anderson, J. R., "Cognitive psychology and its implications," Ewha Womans University Press, pp. 87-115, 2000.
2. Itti, L., Koch, C. and Niebur, E., "Model of saliency-based visual attention for rapid scene analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence(PAMI), pp. 1254-1259, 1998.
3. Choi, K. J. and Lee, Y. B., "Detecting Salient Regions based on Bottom - up Human Visual Attention Characteristic," The Korean Institute of Information Scientists and Engineers(KIISE) : Software and application, Vol. 31, No. 2, pp. 189-202, 2004.
4. Son, J. I., Lee, M. H. and Shin, J. K., "Implementation of saliency map model using independent component analysis," The Korean sensors society, Vol. 10, No. 5, pp. 286-291, 2001.
5. Koch, C. and Ullman, S., "Shifts in Selective Visual Attention : Towards the Underlying Neural Circuitry," Human Neurobiology, Vol. 4, No. 4, pp. 219-227, 1985.
6. Milanese, R., Wechsler, H., Gil, S., Bost, J. and Pun, T., "Integration of Bottom-up and Top-down Cues for Visual Attention Using Non-Linear Relaxation," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 781-785, 1994.
7. Park, S. J., Shin, J. K. and Lee, M. H., "Biologically Inspired Saliency Map Model for Bottom-up Visual Attention," Lecture Notes in Computer Science, Vol. 2525, pp. 418-426, 2002.
8. Ouerhani, N., Bur, A. and Hügli, H., "Linear vs. Nonlinear Feature Combination for Saliency Computation: A Comparison with Human Vision," Lecture Notes in Computer Science, Vol. 4174, pp. 314-323, 2006.
9. Ouerhani, N. and Hügli, H., "MAPS: Multiscale Attention-Based PreSegmentation of Color Images," Lecture Notes in Computer Science, Vol. 2695, pp. 537-549, 2003.
10. Hu, Y., Xie, X., Ma, W. Y., Chia, L. T. and Rajan, D., "Salient Region Detection Using Weighted Feature Maps Based on the Human Visual Attention Model," Lecture Notes in Computer Science, Vol. 3332, pp. 993-1000, 2004.
11. Park, M. C. and Cheoi, K. J., "A Motion - driven Selective Visual Attention System," The Korea Contents Association, Vol. 5, No. 6, pp. 87-96, 2005.
12. Ouerhani, N. and Hügli, H., "A Model of Dynamic Visual Attention for Object Tracking in Natural Image Sequences," Lecture Notes in Computer Science, Vol. 2686, pp. 702-709, 2003.
13. Lee, T. W., Wachtler, T. and Sejnowski, T. J., "Color Opponency is an Efficient Representation of Spectral Properties in Natural Scenes," Vision Research, Vol. 42, Issue 17, pp. 2095-2103, 2002.
14. Kim, T. H. and Jeong, D. S., "An Image Retrieval Technique using Entropy and Color Features," The Korean Institute of Information Scientists and Engineers(KIISE), Vol. 26, No. 3, pp. 282-290, 1999.