

## 적응적인 Saliency map 모델 구현을 통한 얼굴 검출

### Face Detection through Implementation of adaptive Saliency map

김기중, 한영준, 한헌수

서울시 동작구 상도동 511 숭실대학교 전자공학과  
E-mail: {skyainesis, young, hahn}@ssu.ac.kr.

#### 요 약

인간의 시각 시스템은 선택적 주의 집중에 의해 시각 수용체로 도달되는 많은 물체들 중에서 필요한 정보만을 추출하여 원하는 작업을 수행한다. Itti와 Koch는 시각적 주의를 제어할 수 있는, 신경계를 모방한 계산적 모델을 제안하였으나 조명환경에 고정적인 saliency map을 구성하였다. 따라서, 본 논문에서는 영상에서 ROI(region of interest)을 탐지하기 위한 조명환경에 적응적인 saliency map 모델을 구성하는 기법을 제시한다. 변화하는 환경에서 원하는 특징을 부각시키기 위하여 상황에 적응적인 동적 가중치를 부여한다. 동적 가중치는 conspicuity map에 S.K. Chang이 제안한 PIM(Picture Information Measure)을 적용시켜 정보량을 측정 한 후, 이에 따라 정규화된 값을 부여함으로써 구현한다. 제안하는 조명환경에 강인한 적응적인 saliency map 모델 구현의 성능을 얼굴검출 실험을 통하여 검증하였다.

**Key Words** : Saliency Map, Visual Attention, Salient region, Dynamic weights, Face Detection

#### 1. 서 론

인간의 시각 시스템은 선택적 주의 집중에 의해서 시각 수용체로 도달되는 많은 물체들 중에서 필요한 정보만을 추출하여 원하는 작업을 수행한다. 시각적 주의를 제어할 수 있는, 신경계를 모방한 최초의 계산적 모델은 1985년 Koch와 Ullman에 의해서 만들어졌다. 그 후 Itti와 Koch가 작금에 가장 일반화된 saliency map 모델을 제시하였다[1]. 그러나 이 모델들은 시각적 주의에 대한 개념 정립 단계이며, 특정 문제에 국한해서 동작하거나 이론적인 측면에 많이 치중되어 있어 실제적인 문제에 적용하기 어렵다[2]. 따라서, Itti와 Koch가 제안한 모형인 saliency map을 개선시키기 위한 많은 연구들이 진행되었다. Nabil과 그의 동료들은 비선형방식으로 saliency Map을 만들었다[3]. 또한, Nabil과 Yiqun Hu는 각각의 conspicuity map에 일정 가중치를 적용한 saliency Map 모델을 제시하였다[4-5]. 위 saliency Map 모

델은 모든 conspicuity map에 의해 얻어지는 전체 특징값 중 가장 큰 값을 나타내는 값을 취하거나, 각 conspicuity map 특정한 값으로 임계치화 하는 방법을 사용하거나, 적절한 가중치를 두어 조합하는 방법을 사용하였다. conspicuity map에 따라 가중치를 달리하여 saliency map을 사용하였으나 조명환경이나 응용방법에 따라 큰 성능의 차이를 보였다. 따라서 본 논문에서는 ROI(region of interest)를 탐지하기 위해 조명환경에 적응적인 새로운 saliency map 모델을 제안한다. 제안하는 saliency map을 만들기 위하여 색상과 명암도 정보를 기반으로 하여 feature map을 구성한다. conspicuity map은 일반적으로 사용하는 중심 주변 차이 기법으로 만들고, 변화하는 환경의 특징을 부각시키기 위하여 상황에 적응적인 동적 가중치( Adaptive Dynamic Weight )를 부여받는다. 동적 가중치는 conspicuity map에 S.K. Chang이 제안한 엔트로피 측정 기법[6]을 적용시켜 영상의 정보량을 측정 한

후, 이 정보를 이용하여 계산한다. saliency map은 가중치를 부여받은 각각의 conspicuity map들을 조합하여 구성한다. 이렇게 만들어진 saliency map은 영상에서 얼굴 영역을 검출하는 실험을 통하여 그 성능을 검증한다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 본 논문에서 제안하는 saliency Map 모델을 구성하는 방법을, 3장에서는 제안한 모델의 타당성을 검증하기 위해 얼굴 검출에 적용시킨 실험에 대해 언급할 것이다. 마지막으로 4장에서 결론 및 향후 연구에 관하여 논의한다.

## 2. 적응적인 Saliency map 모델

본 논문에서 제안하는 모형은 사전지식 없이 주어지는 시각 자극의 특징에 의해서만 관심 영역을 추출하는 상향식 접근이다. 그림 1은 본 논문에서 제안하는 새로운 saliency map 모델의 구성도를 보여준다.

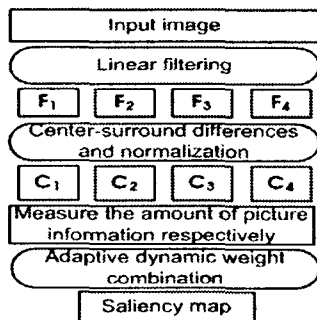


그림 1. Saliency map 구성도

먼저, 제안하는 saliency map 모델은 입력되는 영상들로부터 시각적 주의에 영향을 미치는 특징들로 구성된 여러 개의 feature map을 얻는다. 본 모델에서는 색상 정보와 명암 정보를 특징으로 사용하여 4가지 feature map을 구성한다. 이 두 개의 특징으로 추출해 낸 feature map들을 두드러진 특징을 가지는 conspicuity map들로 변환한다. 그리고 동적 가중치를 구하기 위해, conspicuity map에 S.K. Chang이 제안한 PIM (Picture Information Measure)을 적용하여 각 conspicuity map이 갖는 정보량을 측정한다. 마지막으로 conspicuity map의 중요성에 따라 가중치가 달리 적용되어 saliency

map이 구성된다.

### 2.1 Feature Map

#### 2.1.1 색상 Feature map

색상정보를 뇌로 전달해 주는 인간의 신경회로는 3가지 추상체들에 의한 정보를 '적/녹', '황/청'의 반대 쌍의 색상정보로 바꾸어 전달한다. 입력영상에서 색상 feature map을 얻기 위해 RGB 칼라모델의 적색, 녹색, 청색을 r, g, b라고 할 때 식 (1)에 의해 4개의 색상 채널을 만든다[1].

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

R은 적색 채널, G는 녹색 채널, B는 청색 채널, Y는 황색 채널을 나타낸다. 본 논문에서는 위 4개의 채널을 사용하여 두 개의 색상 Feature map을 식 (2)를 통해 만든다.

$$\begin{aligned} RG(x,y) &= R(x,y) - G(x,y) \\ BY(x,y) &= B(x,y) - Y(x,y) \end{aligned} \quad (2)$$

여기서 RG와 BY는 '적/녹', '황/청'의 반대 쌍의 색상정보를 의미한다.

그리고 얼굴 검출을 위해 다음 식 (3)처럼 피부 색상을 이용하여 Feature map을 만든다.

$$\begin{aligned} I &= 0.6r + 0.27g + 0.32b + 150 \\ Q &= 0.21r + 0.52g + 0.31b + 132 \end{aligned} \quad (3)$$

Skin Region :

$$(163 < I < 199), (123 < Q < 133)$$

#### 2.1.2 명암도 Feature map

국부적인 밝기가 saliency map을 구성하는데 주요기준이 되는 실 영상들이 많으므로 식 (4)에 의해서 명암도 영상을 만들고 이 영상이 Feature map으로 사용한다.

$$I = (r + g + b) / 3 \quad (4)$$

### 2.2 conspicuity map

각각의 Feature map은 주변으로부터 가장 두드러지게 드러나는 부분이 강조되는 특징을 가지게 되는 conspicuity map으로 변환된다. 식 (5)는 conspicuity map을 얻기 위한 Gaussian Pyramid를 수행하는 과정 식을 보여준다.

$$\begin{aligned} P_j(0) &= F_j \\ P_j(i) &= F_j(i-1) * G \end{aligned} \quad (5)$$

각각의 특징을  $j$ 라 하면, 9개의 크기 가우시

안 피라미드  $P_j$ 는 feature map  $F_j$ 에 가우시안 필터  $G(x,y)$ 를 사용하여 점진적으로 subsampling하고 lowpass 필터링하여 얻는다. 여기서  $*$ 는 공간 컨볼루션 연산자이다.

각각의 특징  $j$ 에 대해서, 6개의 중간 다중 크기 conspicuity map들인  $M_{j,k}(1,2,\dots,6)$ 를 식 (6)에 의해서 구한다.

$$\begin{aligned} M_{j,1} &= |P_j(2) - P_j(5)|, & M_{j,2} &= |P_j(2) - P_j(6)| \\ M_{j,3} &= |P_j(3) - P_j(6)|, & M_{j,4} &= |P_j(3) - P_j(7)| \\ M_{j,5} &= |P_j(4) - P_j(7)|, & M_{j,6} &= |P_j(4) - P_j(8)| \end{aligned} \quad (6)$$

6개의 중간 다중 크기 conspicuity map들은 유일한 특징을 가지는 conspicuity map  $C_j$ 로 식 (7)처럼 합산된다.

$$C_j = \sum_{k=1}^6 N(M_{j,k}) \quad (7)$$

여기에서  $N(\cdot)$ 는 정규화 함수이다.

### 2.3 Saliency Map

시각적 주의 모델은 각각의 conspicuity map에 대한 가중치를 결정하고 조합하여 saliency map을 구성한다(식(8)).

$$S = \sum_{i=1}^n w_i c_i \quad (8)$$

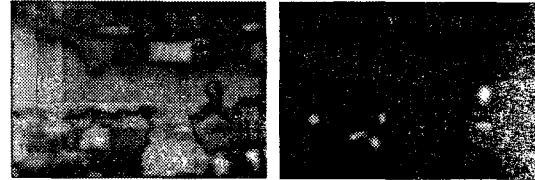
여기서  $n$ 는 conspicuity map의 수이다.

본 논문에서는 saliency map을 구성하기 위해 사용되는 conspicuity map의 중요성에 따라 동적 가중치를 동적으로  $w$ 를 결정한다. 적응적인 동적 가중치(Adaptive Dynamic Weights)  $w$ 는 영상의 엔트로피를 이용하여 구한다. 엔트로피 정량화에는 S-K Chang이 제안한 PIM (Picture Information Measure)을 적용하였다. 본 논문에서 사용한 PIM은 식 (9)로 정의된다.

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

여기서  $L$ 은 양자화 레벨 수,  $h(i)$ 는 각 해당 블록(또는 영상)에서 레벨  $i$ 에 대한 히스토그램을,  $\text{Max}_i(h(i))$ 는 가장 큰 도수의 히스토그램 값을 의미한다. 영상에서 PIM의 평균이 크다는 것은 영상에 엔트로피가 큰 영역이 많다는 것을 의미하고, PIM의 분산이 작다는 것은 영

상에 전체적으로 균일한 크기의 엔트로피 성분이 분포하고 있음을 말해준다. 그림 2는 명암도 영상의 conspicuity map과 피부색상의 conspicuity map을 16×16 사이즈 블록으로 나누어 구한 PIM의 평균과 분산을 도시한 것이다.



(a)명암도 영상의 conspicuity map      (b)피부색상 영상의 conspicuity map  
 그림 2. 영상에서 PIM값들의 평균(M)과 분산(V)

그림 2-(a)는 PIM의 평균이 크지만 분산이 낮으므로 정보량은 많되 그 정보가 전체적으로 퍼져 있음을 알 수 있고 2-(b)는 평균은 낮고, 분산이 크므로 정보량은 적지만 정보가 몰려있다는 것을 알 수 있다. 이러한 PIM의 특성을 이용하여 본 논문에서는 다음과 같은 룰을 만들고 이에 따라 가중치를 적용하였다.

1. 영상의 PIM의 분산이 임의의 임계분산을 넘지 못하면 특징을 갖지 않는다.
2. 영상의 PIM의 평균이 임의의 임계평균을 넘지 못하면 정보량을 가지지 않는다.
3. 영상의 PIM의 평균과 분산이 임계값을 넘기면 정보와 특징을 가진다.

표 1은 가중치 결정 룰을 정리한 것이다.

표 1. PIM을 이용한 가중치 결정 룰

조건	가중치 부여
분산 < $T_v$	0
평균 > $T_m$ && 분산 > $T_v$	$w$
평균 < $T_m$	0

각각의 가중치  $w$ 는 식 (10)에 의해 결정된다.

$$W_{total} = \sum_{j=1}^N \text{Var}_j \quad (10)$$

$$w_i = \text{Var}_j / W_{total}$$

여기서  $N$ 은 선택된 conspicuity map의 수이고,  $\text{Var}_j$ 는 선택된 conspicuity map 각각의 분

산이다.

### 3. 실험 및 결과

제안하는 saliency map 모델의 성능을 분석하기 위하여 실험에 사용한 영상은 인식 및 분석을 하고자 하는 대상에 대한 정보가 주로 존재하는 영상을 대상으로 하였다. 그림 3은 실험 영상에 대한 실험 결과이며, 복잡한 영상에서도 얼굴의 후보영역을 분리해내는 것을 볼 수 있다.

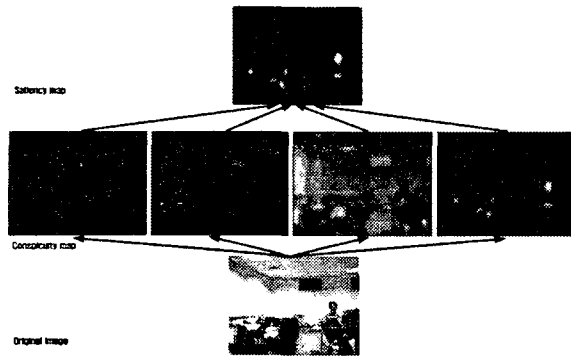


그림 3. 실험영상에서의 얼굴 영역 탐지

실험 결과에서 보듯이 색상 정보와 명암도 정보만을 특징으로 사용하였음에도 불구하고 영상 내에서 지각적으로 중요하다고 여겨지는 주요 영역이 효율적으로 검출되었다.

본 논문은 saliency map으로 분리한 ROI가 얼굴 영역일 가능성이 높다는 가정을 세우고 이 가정을 뒷받침하기 위해 얼굴 비중이 큰 것과 그림 3과 같이 얼굴 비중이 적은 영상 등 여러 상황에 대해서 실험을 진행하였다. 또한 saliency map으로 분리된 ROI에 다른 변환을 가하지 않고 얼굴판단 메커니즘[7]만 적용하여 실험하였다.

표 2는 saliency map으로 분리한 ROI가 얼굴인지 아닌지 판단한 결과를 도시한 것이다.

표 2. saliency map을 이용한 영역 분리 결과

구분	입력개수	판단결과	오차(%)
얼굴영상	100	92	8.6

실험결과는 오차율 8.6%로 얼굴이 들어간 영상에서의 ROI가 얼굴일 수 있다는 가정을 만족할만한 수치로 입증해 주고 있다.

### 4. 결론

본 논문에서는 영상의 엔트로피를 기반으로 한 새로운 saliency map 모델을 제시하였고, 여러 복잡한 영상에 적용하여 그 성능을 테스트하였다. 복잡한 연산을 통한 conspicuity map에 가중치를 부여하여 saliency map을 구성하는 것과 달리 간단한 연산으로도 자연 영상, 얼굴 영상, 여러 복잡한 실험영상에 적용되어 만족할 만한 성능을 보였다. 본 논문에서 제안하는 방법은 ROI가 꼭 얼굴이라는 보장은 할 수 없지만, saliency map의 특징 요소로서 피부 색상을 첨가하였기 때문에 ROI가 얼굴 혹은 피부일 수 있다는 가정을 할 수 있고, 실험 결과는 이 가정을 뒷받침해 주고 있다. 앞으로 더 많은 성능 분석이 필요하겠지만 로봇비전, 감시시스템과 같은 실시간성이 필요한 분야에 응용할 수 있으리라 생각된다.

### 참 고 문 헌

- [1] L. Itti, C.Koch and E. Niebur, "Model of saliency-based visual attention for rapid scene analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence(PAMI), Vol.20, No.11, pp.1254~1259, 1998.
- [2] 최경주, 이일병, "인간의 상향식 시각적 주의 특성에 바탕을 둔 현저한 영역 탐지," 정보과학회 논문지 : 소프트웨어 및 응용, Vol. 31, No. 2, pp.189~202, 2004.
- [3] Nabil Ouerhani, Alexandre Bur, and Heinz Hügli, "Linear vs. Nonlinear Feature Combination for Saliency Computation: A Comparison with Human Vision," Lecture Notes in Computer Science, Vol.4174, pp.314-323, 2006.
- [4] Nabil Ouerhani and Heinz Hügli, "MAPS : Multiscale Attention-Based PreSegmentation of Color Images," Lecture Notes in Computer Science, Vol. 2695, pp.537-549, 2003.
- [5] Yiqun Hu, Xing Xie, Wei-Ying Ma, Liang-Tien Chia and Deepu Rajan, "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.
- [6] 김태희, 정동석 "엔트로피와 색채 특징을 이용한 영상 검색 기법," 정보과학회논문지(A), Vo. 26, No.3, pp.282~290, 1999.
- [7] 임성현 "얼굴색상 및 동작정보의 융합을 통한 블록 기반의 얼굴영역 자동검출기법," 숭실대학교 전자공학과 석사 학위 논문, 2003년 6월