

정규논문 (Regular Paper)
방송공학회논문지 제20권 제4호, 2015년 7월 (JBE Vol. 20, No. 4, July 2015)
<http://dx.doi.org/10.5909/JBE.2015.20.4.557>
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

운동 및 근접 모델을 이용하는 관심맵의 향상

김종인^{a)}, 최창열^{a)}, 김만배^{a)†}

Enhancement of Saliency Map Using Motion and Affinity Model

Jong In Gil^{a)}, Changyeol Choi^{a)}, and Manbae Kim^{a)†}

요 약

정지영상에서 공간 관심맵을 생성하는 다양한 방법들이 소개되어 왔고, 최근에는 동영상의 운동정보를 활용하는 운동 관심맵 예측 기법이 활발히 연구되고 있다. 운동 관심맵은 운동정보 및 영역분할을 활용하고 있지만, 일반적인 영상에서는 만족스러운 데이터를 얻는 것은 어려움이 존재한다. 또한 우수한 관심맵을 얻기 위해서는 객체 운동, 카메라 운동 등의 운동유형 정보가 필요하기 때문에 다양한 자연영상을 대상으로 적용하면 성능 저하가 발생한다. 본 논문에서는 상기 언급한 문제점들을 극복할 수 있는 운동기반 관심맵 생성 방법을 제안한다. 공간 관심맵에 운동 정보를 결합하고, 운동 복잡도를 활용한다. 또한 근접 모델을 이용하여 주변 픽셀들의 관심도를 유사하게 함으로써, 동일 객체 또는 배경 영역이 유사한 값을 가지도록 한다. 실험에서는 다양한 동영상 데이터에 제안 방법을 적용하여 성능 검증을 수행하였다. 공간 관심도의 개선 여부를 증명하기 위해서 공간 관심맵 방법과의 객관적 성능 평가를 통해서 제안 방법이 공간 관심맵보다 운동 픽셀의 경우에 평균적으로 관심도 값이 +38 정도 향상되는 것을 보여준다. 또한 참조 데이터가 있는 4개의 동영상을 대상으로 얻은 ROC는 만족스러운 결과를 보여준다.

Abstract

Over the past decades, a variety of spatial saliency methods have been introduced. Recently, motion saliency has gained much interests, where motion data estimated from an image sequence are utilized. In general, motion saliency requires reliable motion data as well as image segmentation for producing satisfactory saliency map which poses difficulty in most natural images. To overcome this, we propose a motion-based saliency generation that enhances the spatial saliency based on the combination of spatial and motion saliencies as well as motion complexity without the consideration of complex motion classification and image segmentation. Further, an affinity model is integrated for the purpose of connecting close-by pixels with different colors and obtaining a similar saliency. In experiment, we performed the proposed method on eleven test sets. From the objective performance evaluation, we validated that the proposed method produces better result than spatial saliency based on objective evaluation as well as ROC test.

Keyword : Saliency Map, Affinity Model, Motion, Saliency Enhancement

a) 강원대학교 컴퓨터정보통신공학과(Dept. of Computer and Communications Engineering, Kangwon National University)

† Corresponding Author : 김만배(Manbae Kim)

E-mail: manbae@kangwon.ac.kr

Tel: +82-33-250-6395

ORCID: <http://orcid.org/0000-0002-4702-8276>

Manuscript received January 19, 2015; Revised June 29, 2015; Accepted July 17, 2015.

1. 서론

관심맵(saliency map)은 인간이 영상을 인식하면서 눈에 띄는 영역이나 객체를 집중하는 방식을 수학적으로 모델링한 것이다. RGB 영상으로부터 관심맵을 얻는 방법은 그동안 많은 연구가 진행되어 왔다. 관심맵은 영상 리타겟팅(retargeting), 영상 확대 및 축소, 객체 추적, 인식 등에서 활용되는 기술이다. 입력 영상의 통계적 정보를 활용하여 픽셀 값의 빈도수(frequency)가 크거나, 영역이 큰 픽셀에서는 작은 관심도(saliency)를 할당하고, 반대인 경우에는 큰 값을 할당하는 방법이 주를 이룬다. 관심맵 생성 방법은 정지 영상을 대상으로 하여 공간 관심(spatial saliency) 영역을 추출하지만^[1-5], 운동을 이용하는 방법들도 활발히 연구되고 있다^[6-12]. 관심맵 생성은 국부(local) 방식, 전역(global) 방식 또는 두 개의 혼합 방식으로 구분된다.

우수한 관심맵을 얻을 수 있는 전제는 전경객체의 크기는 배경보다는 상대적으로 작아야 하고, 전경객체의 컬러 분포는 배경과 상이해야 한다는 것이다. 이 두 조건을 만족하지 않으면 원하는 관심맵을 얻는 것은 쉽지 않다. 일반적으로 이 조건을 만족하는 영상들은 특수한 경우를 제외하고는 많지 않다. 또한 동일 객체에서도 영역별 컬러가 상이하다면 특정영역만 큰 관심도가 할당되고, 다른 영역에는 작은 관심도가 할당되어, 동일 객체이더라도 상이한 관심맵을 얻는 문제가 종종 발생한다.

본 논문에서는 [4]에서 제안한 공간 관심맵 방법을 기반으로 동영상 운동 정보를 이용하여 개선된 관심맵을 생성하는 방법을 제안한다. 이를 위해서 기존 방법을 지수 함수로 변환한 후에, 운동 정보와 통합한다. 운동 유형으로는 객체 운동, 카메라 운동 등이 있고, 이를 구분하여 활용하는 연구도 있지만^[11], 현실적으로 운동 유형을 판단하는 것은 어려움이 있다. 이를 극복하기 위해서, 운동 유형에 관계없이 적용될 수 있는 운동 복잡도(motion complexity)를 활용한다. 통합 결과는 근접 모델(affinity model)을 이용하여 인접 픽셀들의 관심도를 유사한 값으로 변환한다. 다양한 공간 관심맵 방법들이 소개되었는데, [4]는 전역 방식이고 빠른 처리 속도를 가지면서도 다른 방법들과 비교해서도 성능이 우수한 것으로 알려져 있어서 기본 정지영상 관심도

로 활용한다^[1,2].

대부분의 운동 관심맵 방법들은 영역분할(image segmentation)을 활용한다^[10-11]. 특히 운동 정보의 낮은 신뢰도 때문에 성능 측면에서 효과가 있는 영역분할을 적극 활용하곤 있지만, 동일 객체가 여러 개로 분할영역으로 구성되는 단점이 있다. 관련 연구로 Li 등은 운동정보를 획득한 후에 가우시안 혼합 모델(Gaussian Mixture Model)을 이용하여 배경 및 전경객체의 관심도를 조절하고^[10], Huang 등은 카메라 모션을 제거한 후에 객체의 관심도를 구한다^[11]. 다른 방법으로 Zhai 등은 연속 프레임에서 특징점(feature point)를 찾은 후에 두 영상의 매칭점 및 이동량을 이용하여 관심도를 구한다^[4]. 운동 활용의 다른 문제점은 운동유형의 다양성(예들 들어 객체운동, 카메라 운동)이다. 실제 운동 유형의 분류는 어려운 작업이며, 설사 분류가 얻어지더라도 유형에 따라 분리해서 관심맵을 생성하는 것 또한 복잡한 과정을 필요로 한다. 또한 고속 운동(high motion)에서는 정확도가 많이 낮아지게 된다. 이와는 달리 제안 방법은 배경/전경 분리 등의 영역 분할을 사용하지 않고, 또한 운동 유형에 독립적이기 때문에, 동영상에 쉽게 적용이 가능하고, 또한 근접 모델을 활용하여 주변 픽셀들의 관심도 값을 유사하게 함으로써 영역의 관심도 불일치(discrepancy)를 줄일 수 있도록 한다.

본 논문의 구성은 다음과 같다. 먼저 II장에서는 전체 흐름도를 통해 제안 방법을 자세히 소개한다. III장에서는 수행한 실험 결과를 분석하고, 마지막으로 VI장에서 결론을 맺는다.

II. 제안 방법

그림 1은 제안 방법의 전체 흐름도를 보여준다. RGB 영상에서 관심도(saliency)를 구하고, 두 연속 그레이스케일 영상의 운동 예측에서 얻어진 운동 정보를 이용하여 운동 관심도(motion saliency)를 구한다. 또한 운동 복잡도(motion complexity)를 계산을 한 후에, 이 값을 운동 관심맵에 적용한다. 운동 복잡도는 운동 유형 결정의 낮은 신뢰도를 해결하기 위해서 사용된다. 두 개의 관심도를 통합한

후에 근접 모델(affinity model)을 이용하여 구해진 관심맵의 주변 픽셀들의 값 차이를 감소시킨다.

$$S_i = \sum_{j=1}^N D(I_i, I_j) = \sum_{j=1}^N |I_i - I_j| \quad (1)$$

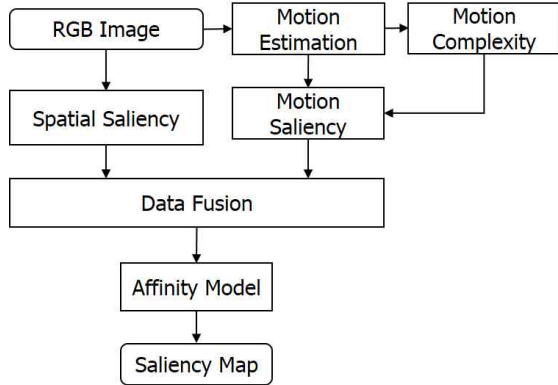


그림 1. 제안하는 운동 관심맵 방법의 흐름도
 Fig. 1. Flow diagram of the proposed motion saliency method

1. 공간 관심도 변환

정지영상의 공간 관심도를 구하는 방법 중의 하나인 Zhai가 제안한 전역 기반 방식은 색상 거리 등을 고려하여 중요도를 계산하는 방식이다^[4]. 픽셀 I_i 의 중요도 값은 입력 영상에서 다른 픽셀들과의 색상 차이(color difference)를 이용하여 계산한다. 픽셀 I_i 의 관심도 S_i 는 다음 식으로부터 얻어진다.

여기서 N 은 영상 I 의 픽셀 개수이다. $D(I_i, I_j)$ 는 컬러 공간에서 두 픽셀의 색상 차이(color distance metric)이다. $I \in \{R, G, B\}$. RGB로 얻어진 값들의 평균 또는 최대값으로 관심도를 구한다.

이 기법을 이용하여 얻은 관심맵은 그림 2에서 보여진다. 그림 2(a)는 원영상이고, 2(b)는 Zhai 방법으로 얻은 관심맵이다. 관심맵 생성은 주로 색의 차이를 이용하기 때문에 배경과 전경의 색의 분포가 유사하면 만족스러운 결과를 얻지 못한다. 또한 같은 객체의 내부의 색이 다르면, 다른 관심도를 가지게 되어 부자연스러운 결과를 얻게 된다.

식 (1)은 두 픽셀 값의 차이를 이용한다. 두 값의 차이가 클수록 더 큰 관심도를 얻는 데 그림 3의 단순 증가 지수 함수를 이용하여 관심도의 차이를 증가하게 할 수 있다. 그림 3에서 수평축은 두 픽셀의 차이 값이고, 수직축은 $[0, 1]$ 로 정규화된 관심도이다. 따라서 식 (1)은 다음과 같이 변환된다.

$$S_i^c = \sum_{j=1}^N \left[1 - \left(1 - \frac{|I_i - I_j|}{255} \right)^2 \right] \quad (2)$$

여기서 I_i, I_j 는 픽셀 i, j 의 컬러 값이다. S_i^c 는 $[0, 1]$ 의 값을 가진다.



그림 2. Zhai의 관심맵^[4]. (a) 입력 영상 및 (b) 결과 관심맵
 Fig. 2. Saliency map obtained by Zhai's method^[4]. (a) input color images and (b) saliency maps

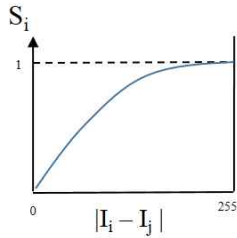


그림 3. 단순 증가 지수함수를 이용하는 관심도 변환
 Fig. 3. Saliency transformation using monotonically-increasing exponential function

2. 근접 모델 (Affinity Model)

기존 방법은 각 픽셀 값만 가지고, 관심맵을 구하는데, 여기에 픽셀간의 공간적 근접성(affinity)을 이용하지 않기 때문에, 두 인접 픽셀의 컬러 값이 다르면, 차이가 큰 관심도를 얻게 되는 문제가 발생한다. 근접 모델은 주로 영역분할이나 필터링에서 사용되어온 기술이다^[13,14]. 주변 픽셀들의 값 차이를 줄여, 가까운 거리이면 유사한 값을 얻게 해주는 필터이다. 일반적으로 지수함수 형태가 주로 이용되는데 두 픽셀 i, j 의 근접도는 식 (3)을 이용하여 얻어진다.

$$\Psi(i, j) = \exp\left(-\frac{\|x_i - x_j\|}{\sigma_a}\right) \quad (3)$$

여기서 x_i, x_j 는 픽셀 i, j 의 좌표이고, σ_a 는 지수함수의 표준편차로 근접도를 조절한다.

식 (3)을 식 (2)에 적용하면 다음의 관심도가 얻어진다.

$$S_i^c = \sum_{j=1}^N \Psi(i, j) \cdot S_{i,j}^c \quad (4)$$

$$= \sum_{j=1}^N \exp\left(-\frac{\|x_i - x_j\|}{\sigma_a}\right) \cdot \left[1 - \left(1 - \frac{|I_i - I_j|}{255}\right)^2\right]$$

이 근접 모델의 효과는 다음 그림에서 보여진다. 그림 4(a)는 Zhai의 관심맵이고, 근접모델을 적용하면 그림 4(b)처럼 전체적으로 주변픽셀의 값이 유사한 값을 가지게 것을 관찰할 수 있다. 가운데 전경객체의 내부가 보다 유사한 값으로 변화되어 더 향상된 관심객체를 관찰할 수 있다.



그림 4. 근접 모델의 효과. (a) Zhai의 관심맵[4] 및 (b) 근접 모델을 적용한 관심맵
 Fig. 4. The effect of affinity model. (a) saliency map [1] and (b) saliency map improved by affinity model

3. Motion 활용

동영상에서 얻을 수 있는 운동정보를 이용하여 관심맵을 구하는 방법은 많은 연구가 진행되어 왔다. 이 중에는 운동을 이용하여 영역분할을 한 후에 영역별 관심도를 구하는 방법이 있다. 정확도가 높으면 좋은 결과를 얻을 수 있지만, 영역분할이 어려운 복잡한 영상에서는 활용에 어려움이 많다. 따라서 영역분할 등의 전처리 과정을 이용하지 않고, 단순히 운동 데이터만을 이용하여 관심도를 향상시키는 방법을 제안한다. 관련 연구로 Li 등은 운동정보를 획득한 후에 가우시안 혼합 모델(Gaussian Mixture Model, GMM)을 이용하여 배경 및 전경객체의 관심도를 조절하고^[10], Huang 등은 카메라 모션을 제거한 후에 객체의 관심도를 구한다^[11]. 또한 Mumtaz 등은 FBM(foreground-background model)을 이용하여 모션 영역분할을 수행한다^[12].

운동유형은 크게 카메라 정지 상태에서 객체만 이동하는 객체운동(object motion)과 카메라가 이동하는 카메라운동(camera motion)으로 구분된다. 카메라운동에는 독립적으로 이동하는 객체가 존재할 수 있지만, 분리의 어려움으로 카메라운동으로 판단한다. 제안 방법은 두 유형을 분리하

지 않고, 운동 데이터만 가지고 관심맵을 구하기 때문에, 복잡도면에서 장점을 가지면서 실제 영상에 구현이 편리하다. 이를 위해서 다음 두 가지를 가정한다; 첫째, 객체운동에서는 운동량이 크면 전경 객체이므로, 관심도는 운동 크기에 비례한다. 또한 운동 영역의 비율은 상대적으로 배경보다는 적다. 그리고, 카메라운동에서는 장면이 카메라와 가까울수록 운동 크기는 증가한다. 따라서 관심도는 운동크기에 비례한다고 할 수 있다. 운동영역은 전체 영상에서 발생하기 때문에, 객체운동보다 운동 영역의 비율이 높다.

운동 벡터(motion vector)는 현재 프레임 I_k 와 이전 프레임 I_{k-1} 으로부터 구해진다. 블록기반 운동예측을 이용하여, 픽셀 또는 블록 단위로 운동벡터 $MV = (MV_x, MV_y)$ 를 추정한다. MV_x 는 수평축으로의 운동 값이고, MV_y 는 수직축으로의 운동 값이다. 운동벡터 크기(magnitude) $|MV|$ 는 다음 식 (5)에서 구해진다.

$$|MV| = \sqrt{MV_x^2 + MV_y^2} \quad (5)$$

관심도가 운동 크기에 비례하기 때문에 이것을 직접 이용하면 다음의 운동 관심도가 얻어진다.

$$S_i^M = [1 - (1 - \frac{\|MV_i\|}{MV_{max}})^2] \quad (6)$$

여기서 MV_{max} 는 운동량의 최대 크기이다.

블록기반 운동예측으로 얻어지는 운동벡터는 신뢰도가 낮기 때문에, 직접 운동벡터를 활용하는 것은 저하된 성능의 관심도를 얻게 된다. 따라서 먼저 운동벡터의 분포를 이용하여, 작으면 객체운동, 높으면 카메라운동으로 분리한다. 운동 분포는 운동 복잡도(motion complexity)에 비례한다. 주목적이 공간 관심도를 향상시키는 것이기 때문에 객체운동이면 객체의 관심도를 증가시키고, 카메라운동이면 운동 특성 때문에, 전체 영상에서 관심도가 증가하기 때문에, 실제 관심도 향상을 얻기가 어렵다. 따라서 운동 복잡도를 이용하여, 적절하게 운동 관심도를 조절하는 것이 필요

하다. 이를 위해서 객체운동에 가까우면 운동의 가중치를 높이고, 반대로 카메라운동에 가까우면 가중치를 낮춘다. 운동 복잡도는 다음과 같이 구해진다.

먼저 두 그레이스케일 영상 Y_k 와 Y_{k-1} 의 차영상 ΔY_k 을 구한다.

$$\Delta Y_k = |Y_k - Y_{k-1}| \quad (7)$$

이 차영상의 평균 μ_k 을 얻은 후에, 이를 이용하여 운동 복잡도인 표준편차 σ_k 를 구한다.

$$\mu_k = \frac{1}{N} \sum_{i=1}^N |Y_k^i - Y_{k-1}^i| \quad (8)$$

$$\sigma_k = \sqrt{\frac{1}{N} \sum_{i=1}^N (\Delta Y_k^i - \mu_k)^2} \quad (9)$$

여기서 Y_k^i 는 Y_k 의 픽셀 i 의 값이다. 운동 복잡도에 따라 운동 관심도의 가중치를 조절하기 위해서 다음 식을 이용한다.

$$\tau = \frac{\sigma_{max} - \sigma}{\sigma_{max}} \quad (10)$$

σ_{max} 는 다양한 실험영상으로 얻은 최대 σ 값을 이용한다. 이 운동 가중치 τ 는 식 (6)의 운동관심도에 다음 식처럼 곱해진다.

$$S_i^M = \tau \cdot [1 - (1 - \frac{\|MV_i\|}{MV_{max}})^2] \quad (11)$$

복잡도 σ 가 크면, τ 는 작아지고, 반대이면 커진다. 카메라운동은 복잡도가 크므로 τ 가 작아지게 되어서 운동 관심도에 낮은 가중치가 적용된다. 따라서 상대적으로 공간 관심도의 영향이 커진다. 반대로 객체운동은 낮은 복잡도로 운동 관심도의 영향이 증가하게 된다. 최종적으로 근접 모델과 운동정보를 이용하여 다음과 같이 관심도를 구한다.

$$\begin{aligned}
 S_i &= \sum_{j=1}^N (S_{i,j}^c + \tau \cdot S_i^M) \cdot \Psi_{i,j} \\
 &= \sum_{j=1}^N [1 - (1 - \frac{|I_i - I_j|}{255})^2] + (\frac{\sigma_{\max} - \sigma}{\sigma_{\max}}) \cdot \\
 &\quad [1 - (1 - \frac{\|MV_i\|}{MV_{\max}})^2] \cdot [-\exp(\frac{\|x_i - x_j\|}{\sigma_a})]
 \end{aligned} \tag{12}$$

여기서 S^c 는 식 (2)의 공간 관심맵, S^M 은 식 (6)의 운동 관심맵, $\Psi_{i,j}$ 는 식 (3)의 근접 모델이다. 운동 관심도에 운동 가중치 τ 를 곱하여, 객체운동에서는 객체의 관심도를 높이고, 카메라 운동에서는 상대적으로 공간 관심도의 비율을 크게 되도록 조절된다. 두 관심맵을 결합한 후에 근접모델을 적용하여 주변 픽셀들 간의 관심도를 유사하게 한다. 위 식을 이용하여 얻은 관심맵에 대한 성능 평가는 다음 장에서 설명한다.

III. 실험 결과 및 분석

제안 방법의 성능 개선을 검증하기 위해서 10개의 데이터 세트를 사용하였다^[11,15-17]. 이중 *redbird*, *horse*, *ski*, *girl*는 [16]의 실험 동영상이고, *birdfall2*는 [15], 그리고 *walking*, *street men*은 [17]의 데이터이다. 이 실험에서는 Zhai의 공간 관심맵과 비교하여 제안된 방법의 성능 비교를 수행하였다. 또한 Huang이 제공한 4개의 실험 동영상 *ball2*, *input2*, *rc2*, *mc2*도 성능 검증에 이용하였다^[11].

그림 5(a)는 동영상에서 얻은 한 장의 영상이고, 5(b)는 운동맵이다. 블록기반 운동예측에서 최대 운동벡터 크기 MV_{\max} 는 40으로 하였다. 5(c)는 Zhai 방법으로 얻은 공간 관심맵이고, 제안 방법에 의해 얻어진 관심맵은 5(d)에서 보여진다. *hen*에서는 전경객체의 관심도가 증가하는 것을 알 수 있다. 두 번째 영상인 *horse*는 상단의 배경이 높은 관심도를 보여주는데, 제안된 방법에서는 이 관심도가 낮아지고 객체에서는 증가하는 것을 알 수 있다. *ski*는 유사한 관심맵을 보이지만, 근접 모델을 이용하기 때문에 주변 픽셀들 간의 값 차이가 적다. *street men*은 사람이 있는 곳에서 관심도가 강해진다. 마지막 *birdfall2*에서는 새가 낙하하는 장면인데, 이 객체의 검출은 어려움이 있어서, 나무가

상대적으로 관심도가 증가한다.



그림 5. 관심맵 결과. (a) 실험영상, (b) 운동맵, (c) Zhai의 관심맵^[4] 및 (d) 제안 방법의 관심맵

Fig. 5. The result of saliency maps. (a) test image, (b) motion map, (c) Zhai's saliency map^[4] and (d) saliency map of the proposed method

제안 방법의 객관적 성능을 평가하기 위해서 다음 성능 평가 방법을 사용한다. 픽셀을 운동 픽셀(motion pixel) 및 정지 픽셀(static pixel)로 분류한 후에, Zhai의 관심맵과 제안 방법의 관심맵을 비교한다. 운동픽셀의 관심도 증가 및 감소, 그리고 정지픽셀의 증가 및 감소 비율을 계산한다. 따라서 운동픽셀의 상승값이 클수록 (반대로 감소값이 작을수록), 반대로 정지픽셀은 증가값이 적으면 (반대로 감소값은 크면), 성능이 개선되었다고 판단한다.

Definitions:

P_M^+ : 운동 픽셀의 관심도가 증가한 픽셀의 비율 (%)

- δ_M^+ : 운동 픽셀의 관심도가 증가한 픽셀의 평균 증가값
- P_M^- : 운동 픽셀의 관심도가 감소한 픽셀의 비율 (%)
- δ_M^- : 운동 픽셀의 관심도가 감소한 픽셀의 평균 감소값
- P_S^+ : 정지 픽셀의 관심도가 증가한 픽셀의 비율 (%)
- δ_S^+ : 정지 픽셀의 관심도가 증가한 픽셀의 증가값
- P_S^- : 정지 픽셀의 관심도가 감소한 픽셀의 비율 (%)
- δ_S^- : 정지 픽셀의 관심도가 감소한 픽셀의 증가값

N_M 은 운동픽셀의 개수이고, N_S 는 정지픽셀의 개수이면, 각 비율과 증가 및 감소값은 아래 식에서 구해진다.

$$P_S^+ = 100 \times \frac{N_S^+}{N} \quad \text{및} \quad P_S^- = 100 \times \frac{N_S^-}{N}$$

$$\delta_M^+ = \frac{1}{N_M} \sum_{i=1, i \in I_M}^N [\text{MAX}(S_{PS}(i) - S_{ZS}(i), 0)]$$

$$\delta_M^- = \frac{1}{N_M} \sum_{i=1, i \in I_M}^N [\text{MIN}(S_{PS}(i) - S_{ZS}(i), 0)] \quad (13)$$

$$\delta_S^+ = \frac{1}{N_S} \sum_{i=1, i \in I_S}^N [\text{MAX}(S_{PS}(i) - S_{ZS}(i), 0)]$$

$$\delta_S^- = \frac{1}{N_S} \sum_{i=1, i \in I_S}^N [\text{MIN}(S_{PS}(i) - S_{ZS}(i), 0)]$$

위 식에서 N_M^+ , N_M^- 는 운동 픽셀에서 각각 값이 증가, 감소한 픽셀의 개수이다. 마찬가지로 N_S^+ , N_S^- 는 정지 픽셀에서 각각 값이 증가, 감소한 픽셀의 개수이다. S_{PS} 는 제안 방법으로 얻어진 관심맵이고, S_{ZS} 는 Zhai의 방법으로 얻어진 관심맵이다.

각 프레임마다 값을 구한 후에 얻은 평균값은 표 1에서 보여진다. 공간 관심도인 Zhai의 방법으로 얻어진 관심맵을 기준으로 제안 방법의 성능을 검증한다. 운동 픽셀의 관심도가 상대적으로 증가하고, 반대로 정지 픽셀의 관심도가 감소한다면 성능이 향상되는 것으로 판단한다.

7개의 실험 동영상에서 운동픽셀 P_M^+ 의 δ_M^+ 의 값이 δ_S^+ 값보다 큰 것을 알 수 있다. redbird, ski, horse, street men, birdfall2, walking, girl은 각각 $(\delta_M^+ - \delta_S^+)$ 의 값이 +34, +13, +30, +54, +63, +60, +21으로 운동픽셀의 관심도가 더 증가하는 것을 알 수 있다. 반면에 δ_S^- 와 비교하면, δ_M^- 의 감소값 $(\delta_M^- - \delta_S^-)$ 은 -20, +2, +3, +6, +23, -22, +12이다. 평균적으로는 운동픽셀 P_M^+ 의 증가값 $(\delta_M^+ - \delta_S^+)$ 은 +38 이고, P_M^- 의 감소값 $(\delta_M^- - \delta_S^-)$ 는 +1이다. 이 결과에 의하면 제안 방법은 Zhai의 방법보다 운동픽셀의 관심도를 전체적으로 증가시키는 것을 알 수 있다.

표 1. Zhai 방법과 제안 방법의 객관적 성능 비교

Table 1. Objective performance comparison of the proposed method with Zhai's method

Test Sequence	Motion pixel	Static pixel		
	$P_M^+ (\delta_M^+)$	$P_M^- (\delta_M^-)$	$P_S^+ (\delta_S^+)$	$P_S^- (\delta_S^-)$
Redbird	61 (+55)	18(-61)	64(+21)	15(-41)
Ski	80(+33)	2(-7)	62(+20)	20(-9)
Horse	71(+48)	8(-37)	72(+18)	7(-40)
Street men	71(+73)	8(-26)	41(+19)	38(-34)
Birdfall2	79(+87)	4(-3)	63(+24)	18(-26)
Walking	79(+81)	21(-44)	67(+21)	32(-22)
Girl	66(+39)	7(-30)	70(+18)	4(-42)
Average	72(+58)	10(-29)	62(+20)	18(-30)

그림 6의 실험 데이터는 수작업으로 만든 그림 6(b)의 ground truth를 포함하고 있다. ball2와 같이 객체의 운동이 명확한 경우에는 개선된 관심맵을 얻을 수 있지만, input2의 경우 상대적으로 객체의 운동이 크기 않고, 느리게 움직이고 있다. 따라서, 운동 정보에 의한 개선은 기대하기 힘들지만, 근접 모델을 사용하였으므로 개선되었음을 확인할 수 있다. mc2의 경우에는, Zhai의 방법에서는 오히려 배경에 높은 관심도가 할당되었고, Seo의 방법에서는 사람의 얼굴을 제외한 대부분의 영역에서 낮은 관심도가 관찰되었다. 그러나 제안 방법을 이용한 결과에서는 배경의 관심도가

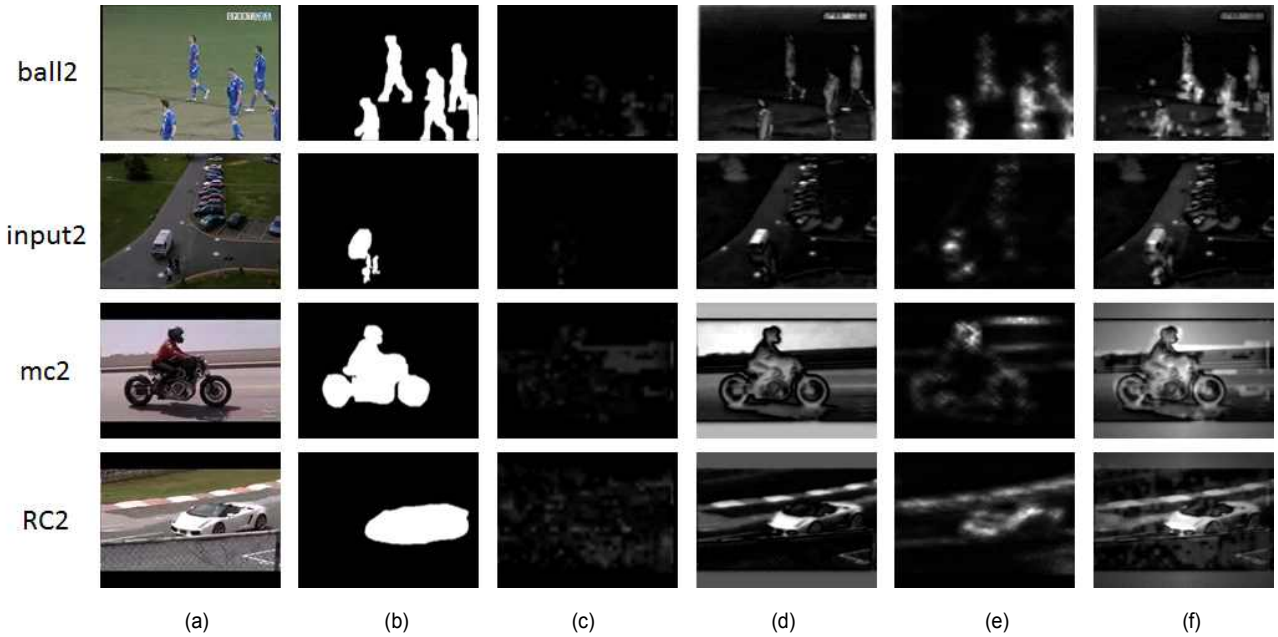


그림 6. Huang의 실험영상^[11]을 이용한 관심맵. (a) 실험영상, (b) ground truth, (c) 운동맵, (d) Zhai의 관심맵^[4], (e) Seo의 관심맵^[17] 및 (f) 제안 방법의 관심맵

Fig. 6. The result of saliency maps from Huang's test sequence^[11]. (a) test image, (b) ground truth, (c) motion map, and saliency map made by (d) Zhai^[4], (e) Seo^[17] and (f) the proposed method

크게 낮아졌고, 사람 및 바이크에 관심도가 증가하였다. RC2에서도 유사한 결과를 보여주고 있다. 전체적으로 객체 운동에서는 우수한 관심맵이 얻어지고, 카메라 운동에서는 효과가 크지 않음을 관찰할 수 있다. 이 관찰은 제안 방법으로부터 예측할 수 있다.

앞서 사용한 평가 방법 이외에 관심맵의 성능 측정을 위

해 가장 널리 사용되는 측정 도구는 ROC이다. ROC는 ground truth와 실험 결과가 얼마나 유사한 지를 나타낼 수 있도록 설계되었다. 관심맵의 임계치를 0~255로 달리하여 측정하면 각각에 대하여 참긍정(True Positive), 거짓 긍정(False Positive), 참부정(True Negative), 거짓부정(False Negative)를 계산할 수 있다. 네 가지 값에 대해 거짓 긍정

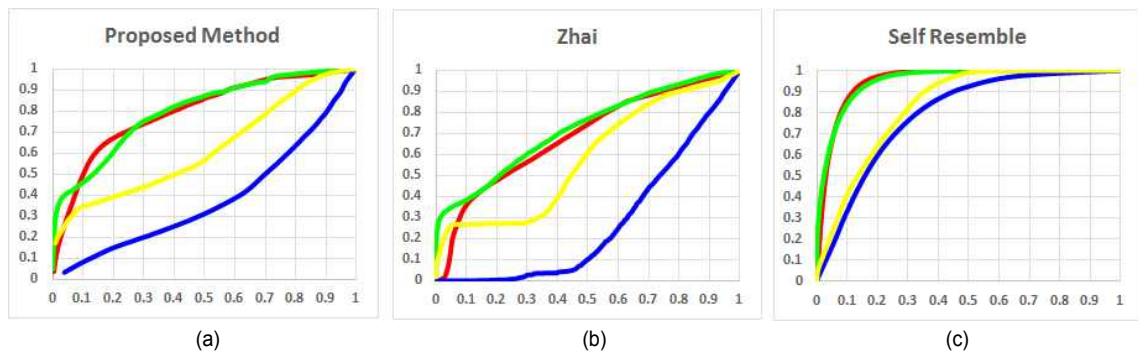


그림 7. 그림 6의 4가지 실험 영상에 대한 ROC 비교. (a) 제안 방법의 ROC 곡선, (b) Zhai의 ROC 곡선 및 (c) Seo의 ROC 곡선
Fig. 7. The ROC comparison using four images of Fig. 6. (a) proposed method's ROC curve, (b) Zhai's ROC curve and (c) Seo's ROC curve

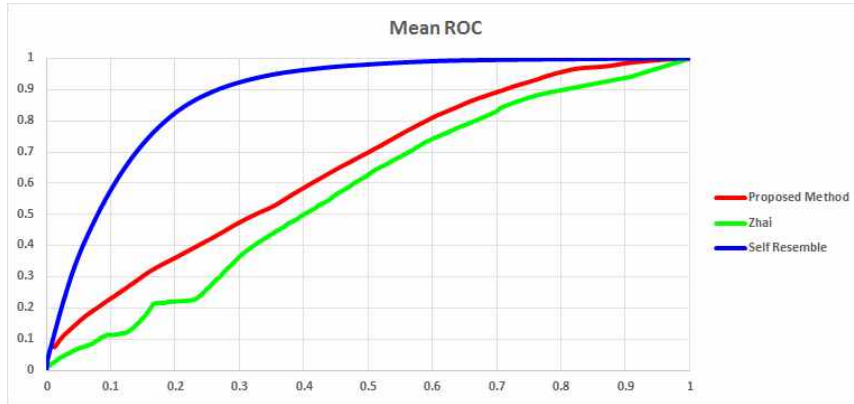


그림 8. 모든 실험 영상에 대한 ROC 곡선의 평균. (제안 방법의 ROC 곡선 평균을 적색으로 표시하였음)
 Fig. 8. The average ROC curve of entire test image. (the mean ROC curve of proposed method is plotted by red color)

률(False Positive Rate)와 참 긍정률(True Positive Rate)를 측정할 수 있다. ROC 곡선은 x축을 거짓 긍정률, y축을 참 긍정률로 하여 도식화한 그래프이다. 그림 7에서 제안 방법과 기존 알고리즘들의 결과로부터 획득한 ROC 곡선을 보여주고 있다. 그림 7, 8의 SelfResemble은 Seo의 알고리즘을 의미한다. 각 방법에 대해 4개의 실험 비디오로부터 측정된 ROC를 나타내었으며, 그림 8에서는 제안 방법 및 기존 알고리즘의 평균 ROC를 측정하여 비교하였다. 알고리즘의 성능 비교는 ROC 곡선의 면적을 비교하여 수행된다. 실험 결과로부터 제안 방법이 Zhai의 알고리즘에 비해 더 큰 ROC 면적을 가지고 있으나, Seo의 방법에서는 더 넓은 ROC 면적을 보여주고 있음을 확인할 수 있다. 그 이유는 Seo의 방법은 대체로 낮은 거짓 긍정률을 보이기 때문이다. 그러나, 그림 6에서 확인하였듯이, Seo의 방법으로부터 획득한 관심맵은 객체의 명확한 윤곽을 확인할 수 없을뿐 아니라, 객체의 내부에 관심도가 낮고, 관심맵이 대체적으로 흐린 단점이 존재한다. 본 논문에서 제안한 방법은 Seo의 방법과 비교하여 높은 거짓 긍정률을 보이고 있는데, 이는 정확하지 못한 운동 정보 추정에 원인을 찾을 수 있다. 본 논문에서는 블록기반 운동예측 기법을 사용하고 있기 때문에 잘못된 운동정보가 나타나게 되면 이는 거짓 긍정률의 증가에 영향을 미치게 된다.

IV. 결론

본 논문에서는 운동정보를 관심맵에 적용하여 기존 공간 영상의 관심도를 향상하는 방법을 제안하였다. 이를 위해서 지수 증가 함수를 이용한 공간 관심도의 향상, 운동유형에 관계없이 운동 복잡도와 운동벡터의 직접 활용, 및 근접 모델링 기법 등을 활용하였다. 제안 방법의 장점은 객체운동, 카메라 운동에 관계없이, 얻은 운동정보를 직접 활용할 수 있고, 이를 위해서 운동 복잡도를 활용하여 적응적으로 운동 관심도의 가중치를 조절할 수 있다. 또한 근접 모델을 이용하여 인접 픽셀들이 유사한 관심도를 가질 수 있도록 하여, 기존 방법들보다 우수한 관심맵을 얻을 수 있다. 실험에서 운동픽셀의 관심도가 비교 방법보다 픽셀당 +38%의 관심도가 증가함을 보여주었다. 제안 방법은 운동유형, 영역 분할 등의 복잡한 과정이 불필요하고, 일반 동영상에 적용될 수 있는 장점을 가지고 있다. 운동 복잡도를 활용하여 객체운동에는 운동관심도의 비중을 높이고, 카메라운동에서는 공간 관심맵의 가중치를 높이도록 하여 적응적으로 관심도를 조절할 수 있다. 따라서 ROC 비교에서는 카메라 운동 영상에서는 상대적으로 성능 저하가 발생할 수 있지만, 객체운동영상에서는 반대로 우수한 성능을 보여준다. 운동 정보를 결합한 관심맵 기법에서는 정확한 운동 정보의 검출이 무엇보다 중요하다. 실험 결과로부터 운동 객

체에 상대적으로 큰 관심도가 부여되는 것을 증명하였으나, 객체의 경계 혹은 내부에 매끄럽지 못한 부분이 발생함을 확인할 수 있었다. 이는 운동 벡터 추정 방법에 대한 내용으로써 독립적으로 연구되어야 할 분야이다. 추후, 운동 검출의 향상을 통해 본 논문에서 제안하는 관심맵 검출 기법의 성능이 향상될 것으로 기대된다.

참 고 문 헌 (References)

- [1] M. Cheng, G. Zhang, N. Mitra, X. Huang and S. Hu, "Global Contrast based Salient Region Detection," IEEE Conf. Computer Vision and Pattern Recognition, pp. 409-416, 2011.
- [2] R. Achanta and S. Susstrunk, "Saliency detection for content-aware image resizing," IEEE Int' Conf. Image Processing, pp.1005-1008, 2009.
- [3] V. Gopalakrishnan, Y. Hu and D. Rajan, "Salient Region Detection by Modeling Distributions of Color and Orientation", IEEE Trans. on Multimedia, Vol. 11, No. 5, Aug. 2009.
- [4] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," Proceedings of the 14th annual ACM Int' Conf. on Multimedia, pp. 815-824, 2006.
- [5] J. Li, Y. Tian, T. Huang and W. Gao "A dataset and evaluation methodology for visual saliency in video," IEEE Int' Conf. on Multimedia and Expo, pp. 442-445, June, 2009.
- [6] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254-1259, Nov. 1998.
- [7] Y. Xia, R. Hu, Z. Huang and Y. Su, "A novel method for generation of motion saliency," Proc. of IEEE 17th Int' Conf. on Image Processing, Sep. 2010.
- [8] Y. Zhu, N. Jacobson, H. Pan, and T. Nguyen, "Motion-decision based spatiotemporal saliency for video sequences," IEEE Int' Conf. on Acoustics, Speech and Signal Processing, pp. 1333-1336, 2011.
- [9] X. Yang, R. Hu, Z. Huang and Y. Su, "A novel method for generation of motion saliency," IEEE Int' Conf. on Image Processing, pp. 4685-4688, Sep. 2010.
- [10] W. Li, H. Chang, K. Lien, H. Chang, and Y. F. Wang. "Exploring Visual and Motion Saliency for Automatic Video Object Extraction," IEEE Trans. on Image Processing, Vol. 22, No. 7, July 2013.
- [11] C. Huang, Y. Chang, Z. Yang and Y. Lin, "Video saliency map detection by dominant camera motion removal," IEEE Tran. on Circuits and Systems for Video Technology, 2012.
- [12] A. Mumtaz, W. Zhang and A. Chan, "Joint motion segmentation and background estimation in dynamic scenes," IEEE Conf. Computer Vision and Pattern Recognition, 2014.
- [13] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 888-905, Aug. 2000.
- [14] Y. Boykov and G. Funka-Lea, "Graph Cuts and Efficient N-D Image Segmentation," Int J. Computer Vision, vol. 70, no. 2, pp. 109-131, 2006.
- [15] D. Tsai, M. Flagg, and J. M. Rehg, "Motion coherent tracking with multi-label MRF optimization," in Proc. Brit. Mach. Vis. Conf., 2010.
- [16] K. Fukuchi, K. Miyazato, A. Kimura, S. Takagi, and J. Yamato, "Saliency-based video segmentation with graph cuts and sequentially updated priors," in Proc. IEEE Int. Conf. Multimedia Expo, pp. 638 - 641, June - July, 2009.
- [17] D. Baltieri, R. Vezzani and R. Cucchiara, "3DPes: 3D People Dataset for Surveillance and Forensics," in Proceedings of the 1st International ACM Workshop on Multimedia access to 3D Human Objects, Scottsdale, Arizona, USA, pp. 59-64, Nov-Dec, 2011. (<http://imagelab.ing.unimore.it/visor/3dpes.asp>)
- [18] H. Seo and P. Milanfar, "Static and space-time visual saliency detection by self-resemblance," Journal of Vision, 9(12): 15, 2009.

저 자 소 개

길 종 인



- 2010년 8월 : 강원대학교 컴퓨터정보통신공학과 학사
- 2012년 8월 : 강원대학교 컴퓨터정보통신공학과 석사
- 2012년 9월 ~ 현재 : 강원대학교 컴퓨터정보통신공학과 박사과정
- 주관심분야 : 3D영상처리, 얼굴 검색, 객체 트래킹, 관심도생성

저 자 소 개



최 창 열

- 1979년 : 경북대학교 전자공학과 학사
- 1981년 : 경북대학교 전자공학과 석사
- 1995년 : 서울대학교 컴퓨터공학과 공학박사
- 1984년 ~ 1996년 : ETRI 컴퓨터연구단 책임연구원 / 연구실장
- 2009년 ~ 2011년 : 강원대학교 IT대학 학장
- 1996년 ~ 현재 : 강원대학교 컴퓨터정보통신공학과 교수
- orcid : <http://orcid.org/0000-0002-8340-4195>
- 주관심분야 : 3D데이터전송, 미디어서비스



김 만 배

- 1983년 : 한양대학교 전자공학과 학사
- 1986년 : University of Washington, Seattle 전기공학과 공학석사
- 1992년 : University of Washington, Seattle 전기공학과 공학박사
- 1992년 ~ 1998년 : 삼성종합기술원 수석연구원
- 1998년 ~ 현재 : 강원대학교 컴퓨터정보통신공학과 교수
- orcid : <http://orcid.org/0000-0002-4702-8276>
- 주관심분야 : 3D영상처리, 모션관심도, 객체인식 및 트래킹