
비디오 재설정 및 3D 압축처리를 위한 어텐션 메커니즘

황재정*

Attentional mechanisms for video retargeting and 3D compressive processing

Jae-Jeong Hwang*

이 논문은 2009년도 한국학술진흥재단의 지원을 받아 수행된 연구임. (KRF-2009-013-D00084)

요 약

이 논문에서는 2D 및 3D 영상의 어텐션량을 측정하여 정지 및 동영상의 재설정 및 압축처리 기법을 제시하였다. 2D 어텐션은 세 개의 주요 구성, 즉, 영상의 세기, 컬러 및 방향성을 고려하였으며, 3D 영상에서 깊이 정보를 고려하였다. 시각적 어텐션은 관심있고 흥미있는 영역이나 객체를 검출하기 위해 희소성을 정량화하는 기법에 의해 구하였다. 왜곡된 스테레오 영상에서 변화된 깊이 정보를 어텐션 확률에 정합시켜서 최종적으로 저위 HVS 반응을 실제 어텐션 확률과 종합하여 스테레오 왜곡 예측기를 설계하였다. 결과로 기존 모델에 비해 효과적인 어텐션 기법을 개발하였으며 이를 비디오 재설정에 적용하여 성능을 입증하였다.

ABSTRACT

In this paper, we presented an attention measurement method in 2D and 3D image/video to be applied for image and video retargeting and compressive processing. 2D attention is derived from the three main components, intensity, color, and orientation, while depth information is added for 3D attention. A rarity-based attention method is presented to obtain more interested region or objects. Displaced depth information is matched to attention probability in distorted stereo images and finally a stereo distortion predictor is designed by integrating low-level HVS responses. As results, more efficient attention scheme is developed from the conventional methods and performance is proved by applying for video retargeting.

키워드

비디오 재설정, 3D 영상, 어텐션, HVS

Key word

Video retargeting, 3D image, Attention, HVS

* 정회원 : 군산대학교 전파공학과(hwang@kunsan.ac.kr)

접수일자 : 2011. 03. 01

심사완료일자 : 2011. 03. 19

I. 서 론

데이터 압축, 시각화 및 디스플레이 기술, 증가된 전송 대역폭 활용, 멀티뷰 및 삼차원 영상 기술이 발전함에 따라 자연영상에 근접한 현실성을 제공하게 되었다. 여기에 힘입어 3-D TV, 3-D 시네마, 3-D 게임, 가상현실 등의 응용분야가 창출되었다 [1-2]. 2-D 영상에 대한 객관적인 화질평가를 구현하기 위해 많은 연구가 진행되었으나, 3-D 영상의 화질평가는 극히 일부의 시도가 있었을 뿐이다. 오늘날 부호화된 영상 시퀀스의 화질평가는 최근 객관화된 메트릭 개발이 보고되고 있음에도 불구하고 [3-4], 아직도 많은 비용이 들고 어려운 주관적 테스트에 의해서 가능한 일이다 [5]. 또한 결과 분석이 일률적이지 않아 신뢰성에 의문이 되고 있다. 결국 3-D 정보를 가장 적절한 방법으로 3-D 영상시스템에 구현하기 위해 현실성 있는 객관적 평가 모델 개발이 필요하다.

현재까지 스테레오 영상에 대한 객관적 평가도구는 깊이정보를 주로 이용하는 방식으로 일부 이루어졌다 [6]. 그러나 깊이정보 계산에는 두 가지 주요 문제가 있다. 첫째, 계산적으로 복잡하고 많은 시간이 소요된다. 둘째, 어떤 깊이가 좋고 나쁜지 판단하기 어렵다는 것이다. 스테레오 비전에 대한 감각은 두 눈의 시점간 차이로부터 나온다. 좌우 영상에서 상응하는 두 점간 거리벡터를 변이라 한다. 최적의 정합점을 찾는 것을 대응문제라 한다. 깊이값은 카메라 기저선과 삼차원 영상점간 거리를 말하며 변이값과 역비례한다.

관측자의 3-D 감각은 디스플레이 기기마다 다르며 여러 종류의 깊이값을 수용할 수 있다. 따라서 깊이 매핑이 동적으로 수행되어 과도한 깊이를 갖지 않도록 할 필요가 있다 [7]. 주파수영역이나 공간영역에서 변이를 검출하는 데 웨스트럼 필터를 이용할 수 있다 [8]. 시각적 감각은 양안간 변이뿐만 아니라 운동시차와 영상현실화의 결과이기도 하다 [9]. 디스플레이 시간 또한 하나의 변수로 고려된다 [10].

어텐션 모델은 하향식(top-down) 과 상향식(bottom-up) 방식을 구분된다. 전자는 물체를 접하기 전에 과거 경험을 전제로 한다. 즉, 인간의 뇌에서 인지적 지식을 바탕으로 한다 [11]. 예를 들어 새가 하늘을 난다고 하자. 관찰자는 새의 다음 행동을 알고 있으며 필력은 날개를 예상한다. 어텐션의 두 번째는 상향식인데, 외부 자극에 의해 구동된다. 즉 색상, 모양, 객체 방향 등 영

상 특징에 대한 인간의 감도를 기반으로 한다.

Milanese [12] 는 상향식 및 하향식 정보가 결합된 어텐션 모델을 제시하였다. 정지영상 및 동영상 모두 적용할 수 있다. 정지영상에서는 주어진 RGB 영상이 먼저 상향식에 의해 분석되어 자극이 현저한 영역을 찾는다. 우선 특징맵 (예, 컬러, 콘트라스트, 곡선, 방향성)을 추출하고 다른 주변 영역과 현저하게 다른 정도를 표현하는 맵을 구성한다. 다음으로 집약과정인데, 각각의 맵을 하나의 맵으로 모은다. 물체 인식 시스템에 의해 하향식 어텐션 정보를 얻고 이를 가미하여 최종 맵을 얻는다.

Chen 등은 각 AO에 관심영역 (ROI: Region of Interest), 어텐션값 (AV: Attention Value), 최소 감지 크기 (MPS: Minimal Perceptible Size) 등 세 속성을 제안하였다 [13]. ROI는 가장 관심있는 물체를 고품질로 부호화하도록 JPEG2000에 표준화되어 있다. 가장 큰 돌출점을 씨앗으로 삼아 ROI를 추출하는 기법도 제안되었다 [14]. ROI는 AO에 대응하는 것으로 간주된다. 서로 다른 AO는 고유의 정보를 갖고 있으며 이를 어텐션값으로 표현하며 영상 내의 중요성을 의미한다. 소형 디스플레이 기기에서 AO 크기가 줄어들면, 더 이상 감지 가능하지 않을 수도 있으며 이를 위해 최소 감지 가능 크기를 정의해야 한다.

Itti는 영상의 어텐션을 계산적으로 구하는 기법을 제안하였다 [15]. 여러 영상 특징을 조합하여 하나의 맵을 구성하고 동적신경망으로 어텐션 위치에 대한 값을 매겼다. 컬러, 세기, 방향성 등 저위 특징으로부터 주요 정보를 구하였다. 검토 가능한 다른 변수로는 스테레오 영상에서 깊이 및 형상 정보이다. 각 정보를 선형적인 중심-주변 시각 반응계에 의해 계산하였는데, 일반적으로 뉴런의 반응은 중심 영역에서 최대로 되는 성질을 이용한다. 중심-주변 모델에서 중심을 세밀하게, 주변은 거칠게 표현하여 특징맵을 구한다. 첫째 특징은 영상의 세기 콘트라스트이며 두 번째는 유사한 방식으로 컬러에 대해 이중 대립계 모델을 이용한다. 중심에서 빨간 색은 여기되는데 초록 색에 대해서는 금지된다. 주변 영역에서는 반대로 작용한다. 세 번째 요소는 방향성 가버 피라미드에 의해 로컬 방향성을 찾는다. 가버 필터는 여현 평면파에 의해 변조된 가우시안 커널 함수로, 기본 시각피질에서 방향선택적 뉴런의 감도를 근사화한다. Li는 Itti 모델로부터 다중스케일 현저성맵을 구하는 계층적 구조를 제안하였다 [16].

시각적 어텐션은 패턴인식, 영상압축, 감시시스템에서 객체 추적, 영상/비디오 요약 등에 응용이 가능하다. [17]에서 Salah 등은 패턴인식을 위한 선택적 어텐션 모델을 제안했다. 수기 문자인식과 얼굴인식에서 좋은 성과를 얻었다. [18]에서 어텐션 위치 즉 ROI 검출을 통해 압축 성능을 향상시켰다. Ma는 움직임 어텐션 모델을 제안했으며 비디오 요약에 이를 이용하였다 [19]. 비디오 신호를 분석하여 모든 가능한 특징을 추출하였다. 세 가지 정보에 대한 어텐션을 종합하기 위해 가중치를 고려한 선형조합법을 사용하였다.

지금까지 살펴 본 바와 같이 2-D, 3-D 공간에서 가장 필수적인 성분을 검출하기 위한 많은 시도가 이루어졌다. 최종 결정 요인으로 인간의 시각 시스템 (HVS: Human Visual System)가 고려되었다. 그러나 대부분의 연구는 HVS 기반 IQA (Image Quality Assessment), 어텐션 모델링, 3-D 깊이 정보 등 다수의 요소 기술 중 하나 또는 일부만을 사용하고 있다. 이 논문에서는 세 기술의 서로 다른 역할을 분석하였으며 비디오 재설정 및 정보량 압축을 위한 어텐션 모델링 기법을 제시하였다.

II. 본 론

2.1. 스테레오 어텐션 모델

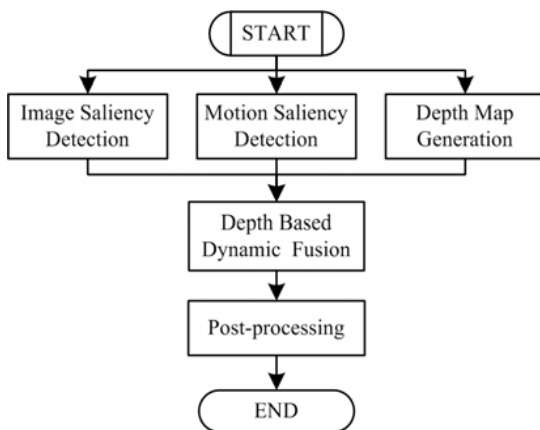


그림 1. 스테레오 어텐션 모델 흐름도.
Fig. 1. Flowchart of the stereo attention model.

어텐션 모델의 일반적 형태는 영상에서 세기, 방향, 컬러 등 세 성분을 대상으로 한다. 비디오에서는 움직임 정보가 추가된다. 3D 비디오는 동일한 물체를 약간 다른 각도에서 관측하게 되며 그만큼 다른 영상을 추출하게 된다. 예를 들어, 인간은 영상에서 빠져 나가는 물체에 더 관심을 갖게 된다. 스테레오 영상에 대한 감각은 따라서 2D 영상과 3D 깊이 정보에 의해 결정될 수 있다 [20]. 깊이는 비디오 객체와 카메라 시스템의 상대적인 거리로 정의된다. 결국 스테레오 영상에 대한 어텐션은 그림 1과 같이 2D 영상의 공간적 현저성(saliency), 움직임 현저성, 그리고 깊이 맵으로 구성되며 다음 식과 같이 정의된다.

$$Att = \{S_s, S_m, D\} \quad (1)$$

여기서 S_s 는 2D 현저성을, S_m 은 움직임 정보를, 그리고 D 는 깊이 정보를 의미한다.

인간이 느끼는 어텐션에서 공간적 특징, 움직임 특징, 깊이 정보가 갖는 중요성은 상황에 따라 다르게 나타날 수 있다. 움직임이 다른 것보다 크게 영향을 미칠 수도 있는데, 예를 들어 강한 움직임이 있는 경우이다. 움직임이 적은 경우에는 공간적 특징이 중요하게 반영되어야 한다. 또한 깊이 정보가 큰 경우에는 이것이 주요 변수로 반영되어야 한다.

움직임 정보의 추출은 전통적으로 픽셀 매칭이나 블록매칭으로 이루어진다. 일반적으로 프레임간 광흐름(optical flow)을 특정하게 된다. 깊이 정보의 추출도 스테레오 매칭이나 그래프 컷 방법에 의해 이루어지며 어텐션 모델을 위해 특별히 고안되는 것보다 독자적인 영역으로 인식된다. 따라서 다음 절에서는 2D 공간 영역에서 효과적인 어텐션 추출 기법에 대해 논한다.

2.2. 제안한 엔트로피 기반 공간 어텐션 모델링

시각적 어텐션은 HVS 기반 평가에서 사용되는 저위 특성들로 구성되지 않는다는 것을 알게 되었다. 시각적 어텐션은 이종적(heterogenous)이거나 동종적(homogeneous), 어둡거나 밝거나, 대칭이거나 비대칭으로 만들어질 수 있다. 이는 HVS 기반 시스템보다 상위 레벨에 해당한다 [21]. 모두 공감하는 하나의 사실은 영상에서 희소성이 있는 정보가 보다 관심을 끌게 되는 것이다. 이를 이용하여 희소성을 양자화하고 모델링하는 연

구를 진행하였다.

자기정보(self-information)는 심볼 확률의 함수로서, 자주 발생하는 심볼은 보다 적은 자기정보량을 다음 식과 같이 갖는다.

$$I(m_i) = -\log(p(m_i)) \quad (2)$$

여기서 $p(m_i)$ 는 발생확률을 의미한다. 발생확률은 영상의 모든 레벨에서 발생빈도를 보여주는 히스토그램으로 표현할 수도 있다.

어텐션점 검출은 인간 어텐션의 최저 레벨을 연고자하는 것으로 외부 자극에 의해 영향을 받는다. 점들간 어떤 관계성을 갖지 않으며 국부적으로 최대 콘트라스트를 갖는 점들을 나타낸다. 임의의 명암 레벨이 주변 화소값으로부터 멀리 떨어져 있으면 눈에 띄게 된다. 두 레벨간 차이가 크면, 현저성도 크게 된다. 임의 레벨 I_k 에 대한 현저성값은 거리맵 D 로부터 계산된다 [11]. 거리맵은 두 레벨간 색차 신호로 정의된다.

$$D(x, y) = \|a_x - a_y\| \quad (3)$$

여기서 a_x 와 a_y 는 [0,255] 범위를 갖는 컬러값이다. 연산자 $\| \cdot \|$ 는 컬러 거리 측정자인데, 연구에서 절대차를 이용하였다. 이렇게 계산된 색차맵을 그림 2에 도시한다.

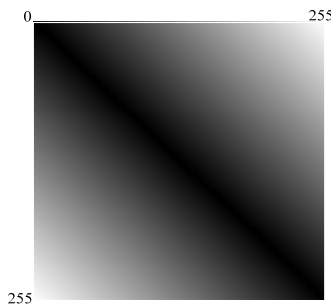


그림 2. 화소값에 대한 거리맵. 밝은 부위는 더 큰 거리값을 갖는다.
Fig. 2. Distance map for pixel values. White level means larger distance.

주어진 영상에서 화소 세기값 a_m 을 알고 있다. 현재 화소값을 $I_k = a_m$ 라 하자. 거리맵은 최대 레벨값으로

정규화되고 히스토그램 $H(m_i)$ 에 곱해져서 다음의 정보 확률을 얻는다.

$$p(m_i) = H(m_i) \times B(m_i) \quad (4)$$

여기서 $H(m_i)$ 는 각 화소의 발생율을 의미하고 화소의 전역 콘트라스트를 의미하는 $B(m_i)$ 는 다음과 같이 정의된다.

$$B(m_i) = 1 - D(m_i) / \max(M) \quad (5)$$

임의 심볼이 다른 것과 현저히 다르면, $B(m_i)$ 는 작게 되며 확률 $p(m_i)$ 도 작게 된다.

모든 영상 화소에 대해 현저성을 계산하는 것보다, 레벨값 $\{a_i, i = 0, \dots, 255\}$ 에 대한 현저성을 계산하여 단순화시킨다. 예를 들어 그림 3에 화소값에 따른 현저성 계산 예를 보인다. 그림 3(c)는 영상의 R 채널에 대한 히스토그램을 왼쪽 그림에 보이고 오른쪽 그림은 해당 현저성값을 표시한 것이다. 수평축은 R 채널 컬러값 $a_m \in [0, 255]$ 을 나타낸다. 자주 발생한다는 것은 중복 정보를 의미하고 따라서 흥미를 덜 유발시키는 인간의 느낌과 유사하다.

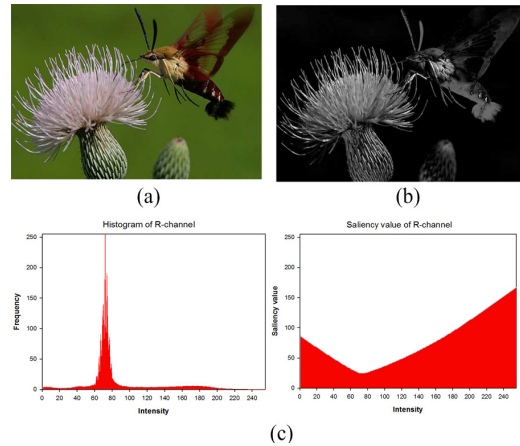


그림 3. 화소단위 현저성 계산 예: (a) 입력영상, (b) 현저성 맵, (c) R 채널에 대한 현저성 값.
Fig. 3. An example of pixel-level saliency calculation. (a) Input image (b) Saliency map (c) Saliency values of R-channel

이렇게 비교적 단순한 알고리즘으로도 출력 현저성 맵 상에 의미있는 객체를 표현할 수 있다. 그러나 영상이 복잡하면 너무 많은 현저성 객체를 표시할 수 있다. 화소의 중요성, 객체의 크기나 형상, 텍스처 정보 등을 분간할 수 없다. 이러한 제약에도 불구하고 이 기법은 가장 현저한 화소값을 얻는 데 유용한 어텐션 화소를 구할 수 있다. 비디오 시퀀스의 재설정을 위해 2D 공간에서 구한 어텐션을 움직임 정보와 깊이 정보를 다음 식과 같이 구하였다.

$$p(m_i) = k_a \cdot p_a(m_i) + k_m \cdot p_m(m_i) + k_d \cdot p_d(m_i) \quad (6)$$

여기서 k_a, k_m, k_d 는 각각 어텐션, 움직임, 깊이에 대한 가중상수이며 $p_x(m_i)$ 는 각각에 대한 확률변수이다.

III. 모의실험 및 결과

제안한 어텐션 모델을 증명하기 위해 정지영상 및 동영상 시험영상으로 이용하였다. 정지영상은 그림 4에 있는 Caps 영상과 그림 5의 Tsukuba 영상이다. Caps 영상은 다섯 개의 모자가 주 어텐션을 갖는 물체이다. 그림 4의 좌하부는 Itti의 어텐션 모델에 의한 결과로 두 개의 모자를 지시하는 부분과 우상부 구름이 주어텐션으로 나타내는데 비해, 우하부의 제안 모델 결과는 전체 모자 부위와 구름 부위가 고르게 나타나고 있다.

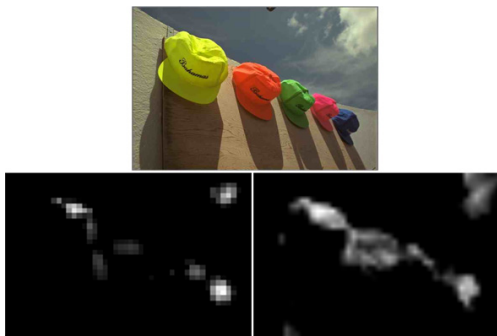


그림 4. 실험영상(Caps)에 대한 어텐션 결과: (상) 원영상, (좌하) Itti 모델 결과, (우하) 제안한 기법 결과.
Fig. 4. Attention results for test image(Caps): (Top) Original image, (Bottom-left) Itti model, and (Bottom-right) proposed model.

스테레오 영상 “Tsukuba.bmp”은 전경 (조각상과 전기스탠드)이 돌출되어 있는 영상이다. 우선 그림 5에 보이는 바와 같이 변이량이 측정되었다. 가장 가까운 객체는 전기스탠드이며 변이맵에서 흰색으로 칠해져 있다. 멀리 있는 배경 (서가)는 반면에 검은 색으로 칠해져 전경으로부터 멀리 떨어져 있음을 나타낸다.

실험용 스테레오 영상은 어텐션 측정 성능을 증명하기 위해 JPEG으로 부호화되었으며, 블록킹, 링잉, 그리고 다른 잡음이 포함되어 있다. 좌우 영상은 소량의 변이가 있고 별도로 부호화되기 때문에 PSNR로 표현한 화질은 약간 다르게 나타난다. 엔트로피 기반 어텐션 모델이 색차신호에 대해 적용되었다. 결과로 명암성분에 대해서는 조각상에 높은 어텐션이 주어지고 색차신호에 대해서는 전기스탠드에 높은 어텐션이 주어짐을 알 수 있다. 각 성분별 어텐션은 종합되어 단일의 어텐션맵으로 구성된다. 방향성맵은 세 번째 요소로서 그림 6과 같이 최종 어텐션 맵에 포함된다. 변이량이 큰 조각상과 전기스탠드에 높은 어텐션을 보임을 알 수 있다.

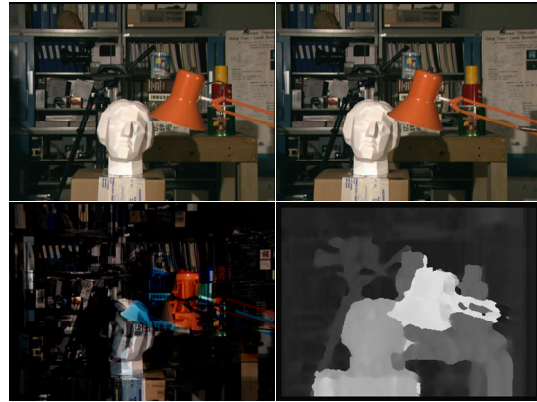


그림 5. 스테레오 영상과 변이맵: (좌상) 좌측 영상 “Tsukuba_r.bmp”, (courtesy of U. of Tsukuba), (우상) 우측 영상, (좌하) 좌우 차분 영상, (우하) 변이맵.
Fig. 5. Stereo images and depth maps: (Left-top) Left image “Tsukuba_r.bmp”, (courtesy of U. of Tsukuba), (Right-top) Right image, (Left-bottom) Left-right difference image, (Right-bottom) Disparity map.

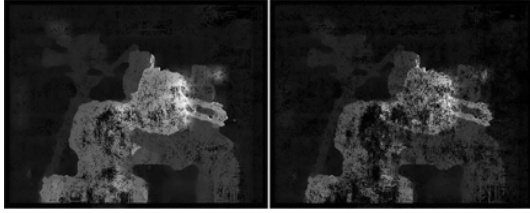


그림 6. 시험 영상(tsukuba)에 대한 어텐션 맵:
(좌) 양자화 레벨 Q0, (우) 양자화 레벨 Q2.
Fig. 6. Attention maps for a test image(tsukuba):
(Left) Quantization level Q0 and (Right) Quantization level Q2.

제안한 기법의 성능을 검증하기 위해 그림 7과 같이 주관적 시험을 실시하여 비교하였다. 비교 대상은 [22]에 제시된 명암영상의 왜곡감도를 컬러영상에 적용한 컬러기반 왜곡감도 (CVDP: Color-based Visible Distortion Predictor)인데 원영상과 왜곡 영상에 대해 어떤 객체가 얼마나 높은 어텐션을 갖는지를 설명해준다. 인간의 시각시스템에서 명암, 주파수, 방향 특성을 고려하여 최종 맵으로 보여준다. 그러나 시각적 어텐션이 고려되지 않은 채 계산되므로 시각적 감도 특성에 기반하여 모든 가능한 왜곡을 보여준다. 제안한 모델 (VAD-SIQAM: Visual Attention & Depth-based Stereo Image Quality Assessment Model)은 시각적으로 어텐션이 높은 영역을 감지 가능 왜곡 영역으로 인지한다. 그림 7에 보이는 바와 같이 CVDP는 주관적 관측결과와 부(-)의 상관성으로 보이는데 비해 제안한 모델은 좌우 영상에서 모두 정(+)의 상관성을 보인다.

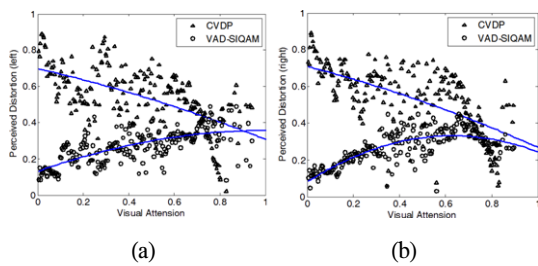


그림 7. 제안한 어텐션 모델(VAD-SIQAM)을 주관적 평가와 비교한 결과: (a) 좌 영상, (b) 우 영상.
Fig. 7. Subjective performance results for the proposed VAD-SIQAM and CVDP: (a) Left image and (b) Right image.

비디오 시퀀스에 대한 시험은 Akiyo와 Football에 대해 이루어졌다. 전자는 한 명의 앵커가 느린 움직임을 포함하고 있으며, Football은 다수의 객체가 빠른 움직임을 갖고 있어 시험영상으로 선정되었다.



그림 8. 시험영상 Akiyo에 대한 재설정 결과:
(좌) 원 영상 프레임, (우상) 제안한 기법에 의해 절반 크기로 재설정된 결과, (우하) 일반적인 재설정 결과.
Fig. 8. Retargeted results for test video Akiyo: (Left) Original frame, (Right-top) The half-size retargeted frame achieved by proposed method, (Right-bottom) Conventional method.



그림 9. 시험영상 Football에 대한 재설정 결과:
(좌) 원 영상 프레임, (우상) 제안한 기법에 의해 절반 크기로 재설정된 결과, (우하) 일반적인 재설정 결과.
Fig. 9. Retargeted results for test video Football: (Left) Original frame, (Right-top) The half-size retargeted frame achieved by proposed method, (Right-bottom) Conventional method.

그림 8에 보이는 바와 같이 일반적인 영상 축소 기법은 가로 세로 양측으로 하향샘플링(decimation)하는 것으로 객체의 크기도 줄어드는 단점이 있다. 이에 반해 제안한 기법에 의한 축소에서는 어텐션이 높은 객체 부위를 중심으로 남기고 의미가 적은 부위를 축소함으로써 원 영상의 내용과 크기를 보존하는 효과가 있다. 그림 9

에 보이는 Football 프레임은 높은 어텐션을 갖는 객체가 물체의 중심에 있지 않고 측면으로 치우친 경우인데, 마찬가지로 제안한 기법은 우하단에 낮은 어텐션 부위를 축소함으로써 중요 객체를 보존하고 있다.

IV. 결 론

이 논문에서는 시각적 어텐션을 추출하는 기법에 대해 논하였다. 어텐션은 2D 및 3D 공간에서 추출될 수 있는데, 2D에서는 영상의 세기, 컬러, 방향을 고려하여 구하는데, 엔트로피 기반 어텐션을 모델링하였다. 3D에 접목하기 위해 움직임 정보와 깊이 정보가 더해졌다. 영상에서 눈에 띄는 것은 새롭게 나타나거나 희소성있는 것이다. 이것을 정량화하여 어텐션을 구하는 데 이용하였다. 세기와 컬러 성분에 대해 희소성 정량화 모델을 적용하였으며 결과를 방향 성분과 조합하였다.

스테레오 영상 각 화소당 변이량을 Zitnick 알고리즘 [23]으로 구하여 어텐션맵을 계산하는 데에 이용한다. 즉 객체의 어텐션은 일반적으로 깊이에 반비례한다. 변이맵을 왜곡된 스테레오 영상의 화질평가에 이용할 때, 변이정보 역시 왜곡되게 되고 원래 깊이보다 깊이변화량이 더 중요하게 된다.

제안한 어텐션 모델은 Itti 모델에 비해 복잡도가 낮으나 중요 객체를 정확히 추출함을 보였으며 인간의 주관적 평가에서도 높은 상관성을 보였다. 어텐션에 기반한 비디오 프레임의 재설정에도 적용하였으며 성능을 입증하였다. 또한 연구 결과는 3-D 영상의 화질평가는 물론 이것을 고려한 3-D 영상부호기 설계 등에 효과적으로 이용 가능하다.

참고문헌

[1] A. Kubota, et al. "Multiview imaging and 3DTV," *IEEE Signal Processing Mag.*, vol.24, no.6, pp.10-21, Nov. 2007.

[2] L. M. J. Meesters, W.A. IJsselsteijn, and P. J. H. Seuntjens, "A survey of perceptual evaluations and requirements of three-dimensional TV," *IEEE Trans. on Circuits and Systems for Video Technol.*, vol. 14,

no. 3, pp. 381-391, Mar. 2004.

[3] M.H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality", *IEEE Trans. on Broadcasting*, vol.50, no. 3, pp.312-322, Sep. 2004.

[4] F. Yang, S. Wan, Q. Xie and H.R. Wu, "No-reference quality assessment for networked video via primary analysis of bit stream", *IEEE Trans. on Circuits and Systems for Video Technology*, 2010.

[5] ITU-T, Recommendation P.910, *Subjective video quality assessment methods for multimedia applications*, April 2008.

[6] S. Narkhede and F. Golshani, "Stereoscopic imaging: a real-time, in depth look," *IEEE Potentials*, vol. 23, no. 1, pp. 38-42, Feb.-Mar. 2004.

[7] G. Sun and N.S. Holliman, "Evaluating methods for controlling depth perception in stereoscopic cinematography", *Stereoscopic Displays and Virtual Reality Systems XX, Proc. of SPIE-IS&T Electronic Imaging*, SPIE, vol. 7237, Jan. 2009.

[8] [Awawdeh03] A. Awawdeh and G. Fan, "Pseudocepstrum for assessing stereo quality of retinal images," *Asilomar Conf. on Signals, Systems and Computers*, vol. 2, pp. 1953-1957, 9-12 Nov. 2003.

[9] M. Ferre, R. Aracil, and M. Sanchez-Uran, "Stereoscopic human interfaces," *IEEE Robotics & Automation Mag.*, vol. 15, no. 4, pp. 50-57, Dec. 2008.

[10] W.A. IJsselsteijn, H. de Ridder, J. Vliegen, "Subjective evaluation of stereoscopic images: effects of camera parameters and display duration," *IEEE Trans. on Circuits and Systems for Video Technol.*, vol. 10, no. 2, pp. 225-233, Mar. 2000.

[11] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," *Proc. the 14th ACM Int. Conf. on Multimedia*, pp. 815-824, Dec. 2006.

[12] R. Milanese, H. Wechsler, S. Gill, J. M. Bots and T. Pun, "Integration of bottom-up and top-down cues for visual attention using non-linear relaxation," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Seattle, pp. 781-785, 1994.

[13] L.Q. Chen, et al., "A visual attention model for

- adapting images on small displays,” *Multimedia Systems*, vol. 9, pp. 353-364, 2003.
- [14] L. Zhang, et al., “Regions of interest extraction based on visual attention model and watershed segmentation,” *10th IEEE Int. Symp. on Multimedia (ISM’08)*, pp. 667-672, 15-17, Dec. 2008.
- [15] L. Itti and C. Koch, “Computational modeling of visual attention,” *Nature Rev. Neuroscience*, vol. 2, no. 11, pp. 194-203, Mar. 2001.
- [16] Q. Li, S. Wang, and X. Zhang, “Hierarchical identification of visually salient image regions,” *Int. Conf. on Audio, Language and Image Process. (ICALIP’08)*, pp. 1708-1712, 7-9 July 2008.
- [17] A.A. Salah, E. Alpaydin, and L. Akarun, “A Selective attention-based method for visual pattern recognition with application to handwritten digit recognition and face recognition,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 420-425, Mar. 2002.
- [18] N. Ouerhani, et al., “Adaptive color image compression based on visual attention,” *11th Int. Conf. on Image Analysis and Process.*, pp. 416-421, 26-28 Sept. 2001.
- [19] Y.-F. Ma, L. Lu, H.-J. Zhang and M. Li, “A user attention model for video summarization,” *Proc. the Tenth ACM Int. Conf. on Multimedia*, pp. 533-542, Dec. 2002.
- [20] Y. Zhang, et al., “Stereoscopic visual attention model for 3D video,” *LNCS*, vol. 5916, pp. 314-324, Dec. 2009.
- [21] M. Mancas, B. Gosselin, and B. Macq, “A three-level computational attention model,” *Proc. of ICVS Workshop on Comput. Attention & Appl.*, 2007.
- [22] S. Daly, “The visible differences predictor: An algorithm for the assessment of image fidelity,” *Digital Image and Human Vision*, Cambridge, MIT press, A. Watson, ed., pp. 179-206, 1993.
- [23] C. Zitnick and T. Kanade, A cooperative algorithm for stereo matching and occlusion detection, *Robotics Institute Tech. Report, CMU-RI-TR-99-35*, Carnegie Mellon University, Oct. 1999.

저자소개



황재정(Jae-Jeong Hwang)

1992 전북대학교 공학박사
1987 군산대학교 전파공학과 교수
2008 호주RMIT대학교 겸임교수
2011 군산대학교 정보전산원장

※ 관심분야: 영상부호화, 디지털 방송, 인터넷 스트리밍, 화질평가, 영상보안