

엔트로피 관점에서 2D 와 3D 동영상의 시각적 정보량 측정방법

안세웅, 이상훈
연세대학교

anse3832@yonsei.ac.kr, slee@yonsei.ac.kr

2D and 3D Visual Information Measurement in terms of Entropy

Sewoong Ahn and Sanghoon Lee
Yonsei University

요 약

최근 2D 와 3D 콘텐츠의 급격한 수요 증가로 인하여 2D 와 3D 공간에서 사람이 인지하는 물체의 시각적 정보량을 정량화할 필요성이 대두되었다. 본 논문에서는 정보이론에 기초하여 엔트로피 관점에서 2D 와 3D 영상의 시각적 정보량을 측정하는 방법을 제시한다. 시각적 정보량을 측정할 때, 기존의 연구에서는 고려되지 않았던 집중영역(saliency), 시각세포의 불균형으로 인한 주변영역 흐림현상인 포비에이션(foveation), 양안합성(binocular fusion)등 인간의 시각적 특성을 반영하였다는 점에서 기존의 연구들과 차이를 둔다. 2D 콘텐츠의 시각적 엔트로피는 단안시에 근거한 질감(texture) 엔트로피와 깊이 엔트로피로 구성되어 있다. 그리고 3D 콘텐츠의 시각적 엔트로피는 2D 에서의 시각적 엔트로피와 양안시에 의한 깊이 엔트로피를 포함한다. 본 논문의 시각적 엔트로피는 2D 와 3D 영상의 시각적 피로도를 측정할 때 사용될 수 있다.

1. 서론

최근 들어 ultra high definition(UHD)와 3D 영상기술의 발전은 소비자의 콘텐츠에 대한 수요를 증가시켰다. 또한, 두 기술의 발전으로 고해상도의 3D 영상의 시청이 가능해져, 소비자는 더욱 생생한 깊이감을 느끼게 된다. 이러한 이유로 공급자는 UHD 와 3D 콘텐츠의 서비스 체계를 소비자 입장에서 미리 준비하는 것이 요구된다[1]. 하지만 아직까지 사람의 인지를 기반으로 한 콘텐츠 정보의 정량화에 대하여 많은 연구가 되지 않은 상태이다. 이 분야를 연구하기 위해 2D 와 3D 콘텐츠의 엔트로피 분석과 정량화를 통하여 시각적 정보를 비교하거나 소비자의 경험적 화질을 예측하는 것이 방법이 될 수 있다. 엔트로피 개념은 콘텐츠 정보 정량화에서 이미 연구되어 있는데 [2], 정보이론에서 엔트로피는 확률 밀도 함수로 변환된 콘텐츠의 불확실성으로 표현된다. 이는 콘텐츠의 정보의 양으로 해석되는데, 이러한 기존의 방식은 사람의 시각적 인지 체계를 반영하고 있지 않다. 따라서 우리는 이 논문에서 Fig. 1 의 시각적 엔트로피 구조도와 같이 동영상의 질감, 단안시 깊이, 양안시 깊이를 이용한 새로운 시각적 엔트로피 요소를 정의한다. 또한, 시각적 정보량을 정확하게 정량화 하기 위하여 우리는 시각 인지체계와 정보이론에 근거한 2D 와 3D 의 시각적 민감도를 연구하고, 콘텐츠에서 각각의 엔트로피가 어떠한 영향을 주는지 분석한다.

2. 시각적 엔트로피 구조

정보이론에 따라 이산 확률 변수 X 와 확률 $p(X) = \{p(x_i) | 1 \leq i \leq n\}$ 에 따른 엔트로피 h 는 다음과 같이 주어진다.

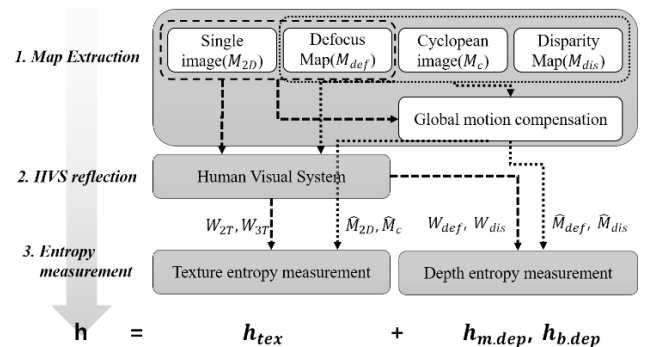


Fig. 1 시각적 엔트로피 구조도

$$h(X) = -\sum_i p(x_i) \log_2^{p(x_i)}. \quad (1)$$

식 (1)을 통하여 사람이 인지하는 3D 와 2D 영상에서의 정보의 양을 정량화 할 수 있는데, 이를 위하여 먼저 영상의 질감과, 깊이에 따른 맵을 추출(map extraction)하여 입력 영상으로 사용한다. 이 입력영상은 먼저 동영상의 엔트로피 측정을 위해 움직임 보상과정(global motion compensation)을 거치게 된다. 그리고 같은 입력 영상으로부터 사람의 시각적 체계(human visual system)를 이용한 2D, 3D 질감 가중치 W_{2T} , W_{3T} 와 초점흐림 가중치 W_{def} , 시차 가중치 W_{dis} 를 결정하게 된다.



Fig. 2 초점흐림과 시차 맵 예시 (a) 'wp1' 158 프레임의 (b) 초점흐림 맵과 (c) 시차 맵 (d) 'mt1' 680 프레임과 그의 (e) 초점흐림 맵 (f), 시차맵

I. 2D, 3D 맵 추출 및 움직임 보상

2D 질감 맵 M_{2D} 은 원본과 동일한 이미지가 사용된다. 깊이 맵은 초점흐림 맵 M_{def} 가 사용되는데, 이는 영상에서 초점으로부터 거리가 멀어짐에 따라 초점의 변화로 인한 흐림 현상이 발생하는 것을 정량화 한 것이다[3]. 3D 질감 맵 M_c 은 왼쪽 영상과 오른쪽 영상을 사람의 인지기반으로 선형 결합한 양안 합성 이미지(cyclopean image)를 사용하게 된다. 이는 왼쪽과 오른쪽 영상의 가보 필터응답으로 비중이 결정되어 합성된다[4]. 양안 깊이 맵 M_{dis} 은 3D 영상에서 왼쪽 오른쪽의 시차(disparity) 맵을 이용한다. 시차는 3 차원에서 깊이를 인지하는 요소가 된다[5]. 2D 와 3D 각각의 맵이 추출되면 [6]에서 연구된 것과 같이 각각의 맵의 모션벡터에서 가장 빈번한 모션 벡터를 전체 모션으로 삼아 두 프레임간의 움직임을 보상하는 맵($\hat{M}_{2D}, \hat{M}_{def}, \hat{M}_c, \hat{M}_{dis}$)을 만든다.

II. 인간 시각 시스템에 근거한 가중치 계산

인간의 시각적 특성을 반영하는 시각적 엔트로피를 측정하기 위하여, 앞선 과정에서 얻은 4 가지 맵의 유형에 따라 각각 시각적 가중치를 계산한다. 질감에 의한 시각적 가중치의 경우 집중영역 감지 알고리즘 [6]을 통해 사람이 집중하고 있는 영역을 찾은 뒤, 포비에이션(foveation)과 합성에 의한 시각적 가중치를 구한다. 포비에이션의 경우는 단안시와 양안시에서 모두 일어나므로, [7]에서 단안시에 의한 포비에이션 함수(f_{ov_b})와 [8]에서 양안시에 의한 포비에이션 함수(f_{ov_m})를 통해 가중치를 구한다. 합성의 경우는 양안시에서만 일어나므로, [9]에서 양안시에 의한 합성 함수(f_{us})를 통해 가중치를 구한다. 위의 함수를 이용한 질감 가중치(W_{2T}, W_{3T})는 다음과 같다.

$$\begin{cases} W_{2T} = f_{ov_m}(M_{2D}) \\ W_{3T} = f_{us}(f_{ov_b}(M_c)) \end{cases} \quad (2)$$

그리고 합성의 경우는 양안시에서만 일어나므로 [9]를 이용하여 가중치를 구한다. 깊이에 의한 가중치는 2D 에 의한 초점흐림 가중치와 3D 에 의한 시차 가중치가 존재한다. 초점흐림에 의한 가중치 W_{def} 는 [3]의 수식을 이용하여 구한다.

$$w_{def}(x, y) = 1 - \frac{FM_{def}(x, y)}{r(V_0 - F)} \quad \text{where } \begin{matrix} 1 \leq x \leq N_w \\ 1 \leq y \leq N_h \end{matrix} \quad (3)$$

F, V_0, r 는 각각 렌즈의 초점거리, 렌즈와 망막의 거리, 렌즈의

반지름을 의미한다. 시차에 의한 가중치 W_{dis} 는 시각 기하 구조를 따른다. W_{def} 와 W_{dis} 의 성분 $w_{def}(x, y)$, $w_{dis}(x, y)$ 는 다음과 같다.

	$h_{tex}(2D)$	$h_{tex}(3D)$	$\alpha h_{m.dep}$	$\beta h_{b.dep}$	h_{2D}	h_{3D}
outdoor	2.84	2.77	2.20	1.46	5.03	6.43
indoor	2.36	2.20	1.84	0.71	4.19	4.75
moving	2.97	2.87	2.11	1.30	5.08	6.27
handheld	2.73	2.56	2.00	1.13	4.73	5.68
static	2.21	2.14	1.98	0.95	4.19	5.07

Table. 1 실험 영상[42]의 시각적 엔트로피의 평균을 외부, 내부, 동적, 정적, 손에 들어진(handheld)로 나누어 분석한 표

$$w_{dis}(x, y) = \begin{cases} \frac{\delta(M_{dis}(x, y))}{\theta_a - \theta_n}, & \text{교차 시차일 경우} \\ \frac{\delta(M_{dis}(x, y))}{\theta_a}, & \text{비교차 시차일 경우} \end{cases} \quad (4)$$

$\delta(X)$ 는 픽셀 시차 X 를 각도시차로 바꾸어 주는 함수이고, θ_a 와 θ_n 은 각각 디스플레이 정면과 후면에서 눈과 초점이 이루는 각도를 의미한다.

III. 시각적 엔트로피 측정

맵 추출 방식과 시각적 가중치 계산을 통해 구한 t 번째 프레임에서 2D와 3D 시각적 엔트로피는 다음과 같다.

$$h_{2D}(t) = \alpha h_{tex}(t) + \beta h_{m.dep}(t). \quad (5)$$

$$h_{3D}(t) = h_{2D}(t) + \gamma h_{b.dep}(t). \quad (6)$$

$h_{2D}(t)$, $h_{3D}(t)$ 는 각각 2D와 3D 영상의 시각적 엔트로피를, $h_{tex}(t)$, $h_{m.dep}(t)$, $h_{b.dep}(t)$ 는 각각 질감, 단안시 깊이, 양안시 깊이 엔트로피를, α, β, γ 는 엔트로피의 상대적인 중요도를 조절하는 지표를 의미한다. 질감 엔트로피 $h_{tex}(t)$ 를 구하는 과정은 다음과 같다. 이미지를 $N \times N$ 블록으로 나누어 이산 코사인 변환을 수행하고, 이를 통해 얻은 $(u, v)^{th}$ 의 계수들을 이용하여 확률질량함수 $p_{u, v}$ 를 얻는다. 정보이론을 이용하여 $(u, v)^{th}$ 이산 코사인 변환 계수의 시각적 엔트로피 $h_{u, v}$ 를 계산한 후, $(u, v)^{th}$ 계수에서의 대비 민감도 함수(contrast sensitivity function) $CSF(u, v)$ 를 이용하여 다음의 수식으로 h_{tex} 를 구한다.

$$h_{tex}(t) = \frac{\sum_{u, v} CSF(u, v) \times h_{u, v}(t)}{\sum_{u, v} CSF(u, v)}. \quad (7)$$

깊이 엔트로피 $h_{x, dep}(t)$ ($X \in \{m, b\}$)를 구하는 과정은 다음과 같다. 3D 공간을 x, y, z 축에 대하여 2^k 구간으로 나누어 총 2^{3k} 개의 3D 블록으로 나눈다. $P_{D/k}(x, y, z)$ 를 k^{th} 레벨에서 $(x, y, z)^{th}$ 3D 블록의 확률을 나타내고, 정보 이론을 이용하여 k^{th} 레벨에서의 깊이 엔트로피 $h_{D/k}(t)$ 를 구할 수 있다. 최종 깊이 엔트로피 $h_{x, dep}(t)$ ($X \in \{m, b\}$)는 $h_{D/k}(t)$ 를 모든 k^{th} 레벨에 대하여 모두 더하여 구할 수 있다. 2D와 3D 영상에서 총 시각적 엔트로피 h_{2D}, h_{3D} 는 각각 t 번째 프레임에서의 시각적 엔트로피 $h_{2D}(t), h_{3D}(t)$ 와 t 번째 프레임과 t-1 번째 프레임과의 차이로 만들어진 시각적

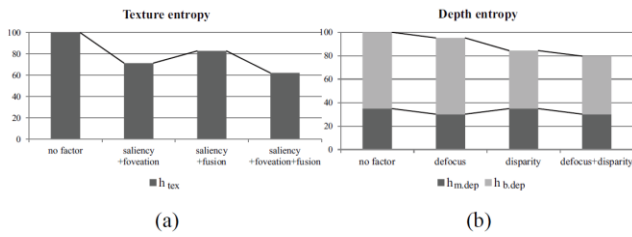


Fig. 3 시각인지 반응이 각각 질감과 깊이의 엔트로피에 미치는 영향 분석 (a) 시각인지 특성별 질감 엔트로피, (b) 시각인지 특성별 깊이 엔트로피

엔트로피 $\Delta h_{2D}(t)$, $\Delta h_{3D}(t)$ 의 합으로 나타낼 수 있다.

$$h_{2D} = \sum_{t \in T} h_{2D}(t) + \sum_{t \in T^c} \Delta h_{2D}(t). \quad (8)$$

$$h_{3D} = \sum_{t \in T} h_{3D}(t) + \sum_{t \in T^c} \Delta h_{3D}(t). \quad (9)$$

여기서 T 는 해당 프레임을 의미하고, T^c 는 해당 프레임을 제외한 프레임을 의미한다.

3. 실험

시각적 엔트로피 구조를 평가하기 위해 IEEE Standards Association Stereoscopic Database [10]의 동영상 사용되었다. 식 (8),(9)를 이용하여 각각 2D 와 3D 에서의 엔트로피를 계산하여 $h_{tex}(2D)$, $h_{tex}(3D)$, $ah_{m,dep}$, $\beta h_{b,dep}$ 의 평균을 계산하였다. 그리고 콘텐츠의 엔트로피간의 관계를 분석하기 위해서 외부, 내부, 동적, 정적, 손에 들어진(handheld) 영상으로 나누어 Table. 1.에 정리하였다. 분석에 의하여 보통 내부에서 찍힌 영상은 제한적인 공간의 인공조형물 때문에 외부에서 찍힌 영상보다 단순한 주파수적 분포를 가지고 있었다. 그리고 외부의 영상은 내부의 영상보다 보통 더 많은 양의 엔트로피를 가진 것으로 나타났다. 게다가 엔트로피는 카메라의 촬영 방법에 따라 많은 영향을 받게 되는데, 카메라의 움직임은 배경의 변화를 가져오기 때문이다. 따라서 엔트로피는 동적 영상, 손에 들어진 영상 정적 영상 순으로 적어지는 현상이 나타났다. 이를 통하여 동영상의 여러 종류에 따른 시각적 엔트로피의 변화를 질감과 깊이를 통하여 관찰할 수 있었다. 또한 우리는 실제로 사람의 시각 인지 체계를 엔트로피 구조에 대입하였을 때, 각각의 시각 인지 모델의 엔트로피만을 이용한 것과 비교하여, 엔트로피와의 연관성을 Fig. 3 에 정리하였다. 이를 통하여 실제로 시각 인지 체계를 반영한 질감 및 깊이 엔트로피가 다른 모델에 비하여 감소하는 결과를 얻을 수 있었고, 이는 이것이 실질적으로 여러 분야에 응용될 수 있는 요소가 됨을 확인할 수 있었다.

4. 결론

동영상 콘텐츠 서비스에서 시각적 정보의 양을 인간의 시각 시스템에 기반하여 측정하는 것은 매우 중요하다. 우리는 이 연구를 통하여 2D 와 3D 동영상의 엔트로피를 측정하는 새로운 방법을 고안하였고, 질감과 깊이 정보를 통해 정확한 측정이 가능하였다. 2D 콘텐츠의 질감과 깊이에 대한 시각적 엔트로피를 분석하는 것을 넘어서 3D 시각적 엔트로피를

분석하는 것은 3D 콘텐츠의 통계적 분산이나 무작위성을 정량화 하는 역할을 하며, 무작위성이나 역동성, 복잡성을 분석하는 것은 3D 콘텐츠를 이해하는데 중요한 요소가 된다. 이 두 가지 시각적 엔트로피의 동영상적 특성을 이용하여 콘텐츠 별로 다양한 적용이 가능한데, 시각적 화질평가나, 압축전송, 집중영역 예측 알고리즘에 사용될 수 있다. 앞으로 우리의 연구는 시각적 엔트로피를 기반으로 다양한 응용에 적용할 것 이다.

5. 감사의 글

본 연구는 미래창조과학부가 지원한 2015년 정보통신·방송 (ICT) 연구개발사업의 연구결과로 수행되었음.

6. 참고문헌

[1] F.-L. Luo, W. Williams, R. M. Rao, R. Narasimha and M. J.Montpetit, "Trends in signal processing applications and industry technology," IEEE Signal Process. Mag., vol. 29, no. 1, pp. 184, Jan. 2012.

[2] Ihara, Shunsuke (1993). Information theory for continuous systems. World Scientific. p. 2. ISBN 978-981-02-0985-8.

[3] A. P. Pentland, "A new sense for depth of field," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-9, no. 4, pp. 523-531, Jul. 1987.

[4] M. J. Chen, C. C. Su, D. L. Kwon, L. K. Cormack, and A. C. Bovik, "Full-Reference Quality Assessment of Stereopairs Accounting for Rivalry," Asilomar Conf. on Signals, Syst. And Computers, Nov. 2012.

[5] M. S. Banks, S. Gepshtein and M. S. Landy, "Why is spatial stereoresolution so low?," J. Neurosci., vol. 24, no. 9, pp. 2077-2089, Mar. 2004.

[6] H. Kim, S. Lee and A. C. Bovik, "Saliency Measurement on Stereoscopic Videos," IEEE Trans. Image Process., vol. 23, no. 4, pp. 1476- 1490, Apr. 2014.

[7] K. Lee, A. K. Moorthy, S. Lee and A. C. Bovik, "3D Visual Activity Assessment based on Natural Scene Statistics," IEEE Trans. Image Processing, vol. 23, no. 1, pp. 450-465, Jan. 2014.

[8] S. Lee, M. S. Pattichis and A. C. Bovik, "Foveated video compression with optimal rate control," IEEE Trans. Image Process., vol. 10, no. 7, pp. 977-992, Jul. 2001.

[9] T. Ohshima, H. Yamamoto and H. Tamura, "Gaze-Directed Adaptive Rendering for Interacting with Virtual Space," Proc. IEEE Virtual Reality Ann. Int'l Symp. (VRAIS), Santa Clara, CA, pp. 103-110, 1996.

[10] (2008). IEEE Standards Association Stereoscopic Database [Online]. Available: <http://grouper.ieee.org/groups/3dhf/>