

CI-002 顕著性マップに基づく 3DCG 画像の関心領域推定と視覚的注意に関する考察

Region Of Interest (ROI) Estimation and Visual Attention of 3D CG Images Based on Saliency Map

河畑 則文[†]

Norifumi Kawabata

1 まえがき

画像や映像を視聴する際には、ヒトは、何かしらの重要そうだと想定される部分かどうかを自分で判断して、効率的に情報を獲得している。このヒトの大変高度な知能をコンピュータ上で再現することができないかということで、今までに、コンピュータ画像処理を用いて、重要度に応じて画像情報を選択するような研究がなされてきた [1, 2]。

一方で、4K, 8K 放送を始めとする画像・映像の高精細化、高品質化に伴い、コンテンツ制作者側としては、ユーザがどこに関心があるのか、どのようなパターンに関心があるのかということ事前に把握した上で、コンテンツを作成することが大変重要である。実際に、コンテンツ制作者が意図して作成したとしても、必ずしもユーザが意図しているものと一致しているとは限らない。また、ヒトによる主観的な評価とコンピュータによる客観的な評価でも異なる。それゆえに、視覚的注意の観点から、どこに着目しているのかという視覚情報を事前に取得可能な画像情報システムの開発とその検証をしていく必要がある。

本論文では、3DCG 画像の物体領域、背景領域のどこに関心領域があるかを顕著性マップにおける確率的モデルの観点から推定し、定量的な評価及び考察を行った。

2 関連研究

本節では、関連研究に関して、(1) 顕著性マップ、(2) 視覚的注意、(3) 画質評価、の観点から述べていく。

顕著性マップについては、文献 [3] では、Spectral Residual と呼ばれる方法を提案している。この手法では、2次元で表現される画像データをフーリエ空間（周波数に変換・表現した空間）で処理を行うことで、特徴を捉えている。文献 [4] では、顕著性メカニズムから派生したモバイルプラットフォームと新規特徴量を使うことにより人物検出の検討を行っている。また、300fps 物体推定に関する 2 値ノルム勾配 [5]、顕著領域検出に基づくグローバルコントラスト [6]、画像データベースにおける集合的な顕著性に

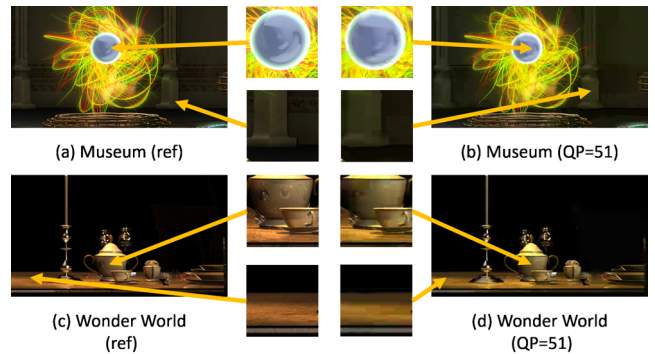


図 1: 実験で使用した 3DCG 静止画像
 (“Museum” (上) “WonderWorld” (下))

関するモデル [7]、ソフトな画像抽象を用いた効果的な顕著領域の特定 [8] が挙げられる。

視覚的注意については、局所的カラーの対称性 [9]、視覚的感覚と品質評価 [10]、ボトムアップモデルに対するアプローチ [11]、基本図形である球 [12]、立体視画像 [13]、多視点画像に関する立体視の視差制御指導型視覚的注意 [14]、視覚的注意に基づくテクスチャのメトリック手法 [15] の検討がされている。

画質評価については、主観的な視覚的評価を用いた立体視画像の視覚的注意に関する評価 [16]、Reduced Reference (RR) 型画質評価に基づく視覚的注意コンテンツの劣化 [17]、事前注意と空間周波数特性を考慮した No Reference (NR) 型画質評価 [18]、アテンションドリブン型の NR 型画質評価 [19] の検討がされている。

3 実験

本研究で用いた 3DCG 画像コンテンツは、NICT が無償で提供している図 1 のような 3DCG コンテンツ (Museum (M), Wonder World (W)) である。画像の生成に関しては、まず、8 視点分の CG カメラを構築し、カメラワーク、レンダリングを行い、8 視点分の静止画像を生成した。本コンテンツは、本来、多視点 3D 画像コンテンツであるが、本研究では、8 視点分の静止画像のうち、1 視点分の単視点画像を用いた。

[†] 東京理科大学 理工学部
Faculty of Science and Technology, Tokyo University of Science

顕著性マップ (saliency map) とは、ヒトが視覚的注意の観点から画像や映像中のどの部分を注視し易いのかを、画像・映像の各ピクセルの顕著性を計算したものである。顕著性マップはグレースケールで表現することもあるが、ヒートマップのようにカラーで表現することもある。本研究では、グレースケールで顕著性マップとして表現していく。OpenCV 3.3.1 に組み込まれているライブラリを用いることにより、顕著性マップの実装を Python 3.6.13 で行った。Spectral Residual Model は以下の式 (1) で表すことができる。そして、OpenCV 3.3.1 のライブラリでは、StaticSaliencySpectralResidual_create() を用いて実装を行うことができる。

$$H(\text{Image}) = H(\text{Innovation}) + H(\text{Prior Knowledge}) \quad (1)$$

次に、spectral residual から顕著性マップに以下の式 (2) のように変換することを考える (SR)。

$$H(R(f)) = H(L(f)|A(f)) \quad (2)$$

また、強度図 (intensity map) を作成するために、OpenCV 3.3.1 では、StaticSaliencyFineGrained_create() を用いて実装することができる (FG)。

ヒトは日常的にどこかへあるいは何かに注意を向けることがあり、それは脳内の過程が存在するといわれている。視覚的注意 (visual attention) については特定の定義がなされていない [20]。現在では、注意にかかわる多くの興味深い現象が知られている。例えば、空間的注意や注意の効果、オブジェクトへの注意、視覚探索と結合錯誤、生理的な側面など様々ある。本研究では画像中の視覚的注意というものを考えているため、空間的注意や注意の効果、何かしらのオブジェクトへの注意が適切であろうと推測できる。

実験方法としては、統合開発環境 Anaconda, Python 3.6.13, OpenCV 3.3.1 を用いて、顕著性マップのアルゴリズムを構成し、プログラムとして動作させた。読み込んだ画像は H.265/HEVC で符号化された 3DCG 画像を用いた。量子化パラメータは $Q = ref, 20, 25, 30, 35, 40, 51$ の 7 種類を設定し実験を行った。評価方法としては、原画像と評価画像を比較して目視で顕著性を測定できているかを見る。評定者による評価実験により主観的な評価は今後の課題になるが、本研究では客観的指標として、どの程度、顕著性が測れるのかということに着目した。

4 実験結果と考察

実験で実装したプログラムを動作させて得られた Spectral Residual の実験結果の画像を図 2～15 に示す。また、実



図 2: “Museum” の SR ($Q = ref$)

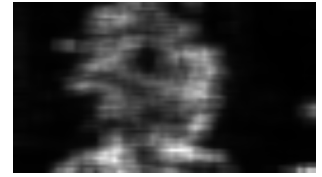


図 3: “Museum” の SR ($Q = 20$)



図 4: “Museum” の SR ($Q = 25$)



図 5: “Museum” の SR ($Q = 30$)



図 6: “Museum” の SR ($Q = 35$)

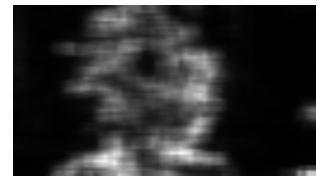


図 7: “Museum” の SR ($Q = 40$)



図 8: “Museum” の SR ($Q = 51$)

験で実装したプログラムを動作させて得られた Fine Grained の実験結果の画像を図 16～29 に示す。

実験結果から、目視の観点からは、“Museum” については、Spectral Residual (SR) では、 Q に関係なく、物体の境界線の枠における顕著性は確認できる。一方、物体の中にある球に関しては、真っ黒になっており、顕著性が見られないと判断された。Fine Grained (FG) では、“Museum” の物体部分の区別はつかないが、背景部分は領域分割されている。“WonderWorld” については、SR では、 Q に関係なく、質感のある部分に関して顕著性が見られた。一方、背景部分に関しては顕著性が見られなかった。FG では、符号化劣化により椅子の部分が欠落してしまっているが、SR の場合よりは見分けがつく。このように、SR と FG の両方の側面から顕著性マップを目視していくことにより、物体部分と背景部分の顕著度の違いや特徴を把握することが可能となった。

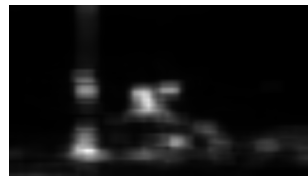
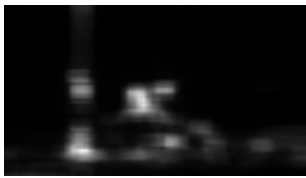


図 9: “WonderWorld” の
SR ($Q = ref$)

図 10: “WonderWorld” の
SR ($Q = 20$)

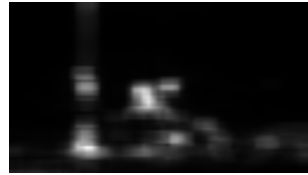
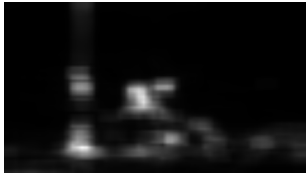


図 11: “WonderWorld” の
SR ($Q = 25$)

図 12: “WonderWorld” の
SR ($Q = 30$)

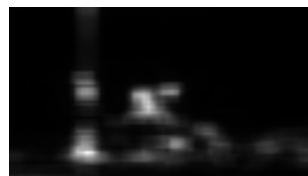


図 13: “WonderWorld” の
SR ($Q = 35$)

図 14: “WonderWorld” の
SR ($Q = 40$)

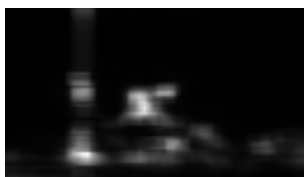


図 15: “WonderWorld” の
SR ($Q = 51$)



図 16: “Museum” の
FG ($Q = ref$)

図 17: “Museum” の
FG ($Q = 20$)

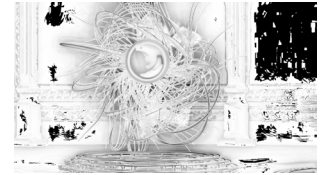


図 18: “Museum” の
FG ($Q = 25$)

図 19: “Museum” の
FG ($Q = 30$)



図 20: “Museum” の
FG ($Q = 35$)

図 21: “Museum” の
FG ($Q = 40$)



図 22: “Museum” の
FG ($Q = 51$)

5 まとめ

本研究の結果から、今回の場合は、CG コンテンツの依存はそれほど見られず、SR と FG の両方を活用することにより顕著性の特定をすることが可能ではないかと推測される。今後の展望として、バウンディングボックスによる位置特定、テクスチャ解析による数値化による定量的評価、主観評価実験や注視点評価も考慮した総合的な顕著性マップと視覚的注意に関する考察を行えば、より詳細に状況把握ができると期待される。

参考文献

- [1] N. Kawabata and M. Miyao: “3D CG Image Quality Metrics by Regions with 8 Viewpoints Parallax Barrier Method,” *IEICE Trans. on Fundamentals*, Vol.E98-A, No.08, pp.1696–1708, August 2015.
- [2] N. Kawabata and M. Miyao: “Multi-view 3D CG Image Quality Assessment for Contrast Enhancement Based on S-CIELAB Color Space,” *IEICE Trans. on Inf. and Syst.*, Vol.E100-D, No.07, pp.1448–1462, July 2017.
- [3] X. Hou and L. Zhang: “Saliency Detection: A Spectral Residual Approach,” *IEEE CVPR2007*, 8 pages, June 2007.
- [4] S. Montabone and A. Soto: “Human Detection Using a Mobile Platform and Novel Features Derived From a Visual Saliency Mechanism,” *Image and Vision Computing*, Vol.28, No.3, pp.391–402, March 2010.
- [5] M. Cheng, Y. Liu, W. Lin, Z. Zhang, P. L. Rosin, and P. H. S. Torr: “BING: Binarized normed gradients for objectness estimation at 300fps,” *Computational Visual Media*, Vol.5, No.1, pp.3–20, March 2019.



図 23: “WonderWorld” の
FG ($Q = ref$)



図 24: “WonderWorld” の
FG ($Q = 20$)



図 25: “WonderWorld” の
FG ($Q = 25$)



図 26: “WonderWorld” の
FG ($Q = 30$)



図 27: “WonderWorld” の
FG ($Q = 35$)



図 28: “WonderWorld” の
FG ($Q = 40$)



図 29: “WonderWorld” の
FG ($Q = 51$)

- [6] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S. Hu: “Global contrast based salient region detection,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.37, No.3, pp.569–582, March 2015.
- [7] M. Cheng, N. J. Mitra, X. Huang and S. Hu: “SalientShape: group saliency in image collections,” *The Visual Computer*, vol.30, pp.443–453, 2014.
- [8] M. Cheng, J. Warrell, W. Lin, S. Zheng, V. Vineet, and N. Crook: “Efficient Salient Region Detection with Soft Image Abstraction” *IEEE ICCV2013*, 8 pages, 2013.
- [9] G. Heidemann: “Focus-of-Attention from Local Color Symmetries,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.26, No.7, pp.817–830, July 2004.
- [10] Z. Lu, W. Lin, X. Yang, E. Ong, and S. Yao: “Modeling Visual Attention’s Modulatory Aftereffects on Visual

Sensitivity and Quality Evaluation,” *IEEE Trans. on Image Processing*, Vol.14, No.11, pp.1928–1942, November 2005.

- [11] O. L. Meur, P. L. Callet, D. Barba, and D. Thoreau: “A Coherent Computational Approach to Model Bottom-Up Visual Attention,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.28, No.5, pp.802–817, May 2006.
- [12] I. Bogdanova, A. Bur, and H. Hügli: “Visual Attention on the Sphere,” *IEEE Trans. on Image Processing*, Vol.17, No.11, pp.2000–2014, November 2008.
- [13] Y. Fang, J. Wang, M. Narwaria, P. L. Callet, and W. Lin: “Saliency Detection for Stereoscopic Images,” *IEEE Trans. on Image Processing*, Vol.23, No.6, pp.2625–2636, June 2014.
- [14] J. Lei, S. Li, B. Wang, K. Fan, and C. Hou: “Stereoscopic Visual Attention Guided Disparity Control for Multiview Images,” *J. Display Technology*, Vol.10, No.5, pp.373–379, May 2014.
- [15] S. Varadarajan and L. J. Karam: “A No-Reference Texture Regularity Metric Based on Visual Saliency,” *IEEE Trans. on Image Processing*, Vol.24, No.9, pp.2784–2796, September 2015.
- [16] H. Kim and S. Lee: “Transition of Visual Attention Assessment in Stereoscopic Images With Evaluation of Subjective Visual Quality and Discomfort,” *IEEE Trans. on Multimedia*, Vol.17, No.12, pp.2198–2209, December 2015.
- [17] J. Wu, Y. Liu, L. Li, and G. Shi: “Attended Visual Content Degradation Based Reduced Reference Image Quality Assessment,” *IEEE Access*, Vol.6, pp.12493–12504, March 2018.
- [18] L. Liu, T. Wang, and H. Huang: “Pre-Attention and Spatial Dependency Driven No-Reference Image Quality Assessment,” *IEEE Trans. on Multimedia*, Vol.21, No.9, pp.2305–2318, September 2019.
- [19] D. Chen, Y. Wang, and W. Gao: “No-Reference Image Quality Assessment: An Attention Driven Approach,” *IEEE Trans. on Image Processing*, Vol.29, pp.6496–6506, May 2020.
- [20] 塩入論, “視覚的注意—現象とそのメカニズム—”, *光学*, Vol.28, No.5, pp.261–272, 1999.