

目と手の協応運動における視覚的顕著性の役割 —人工網膜への応用にむけて—

The Role of Visual Saliency Model in Coordinative Eye and Hand Movement -Towards Applications to Retinal Prostheses-

○ 神田 嵩臣 (大阪大) 神田 寛行 (大阪大)
正 長井 志江 (大阪大) 不二門 尚 (大阪大)
正 浅田 稔 (大阪大)

Takaomi KANDA, Osaka University, takaomi.kanda@ams.eng.osaka-u.ac.jp
Hiroyuki KANDA, Osaka University
Yukie NAGAI, Osaka University
Takashi FUJIKADO, Osaka University
Minoru ASADA, Osaka University

Sight is important for our motion generation such as walking and reaching for an object. However, it is not evident which visual features give helpful information for motion generation. In this paper, we examine the prior visual features used for coordinative eye and hand movement. Our system is aiming at applications to retinal prostheses. Subjects participating in our experiment were asked to look at a low-resolution image converted from an image captured by a head-mounted camera and to reach for a target. We compared the accuracy of reaching between two different converters: the visual saliency model and only brightness feature. The experimental results showed that the saliency map retrieved important features and gave robust sight against changing environments.

Key Words: Visual attention model, Motion generation, Retinal prosthesis

1. はじめに

視覚は人間が生活する上で重要な感覚器であり、視覚から得られる情報量は全体の 83% を占めるといわれている[1]。そのため、人間の歩行や物体把持といった行動生成にも大きく関与しているが、視覚から得られる多くの情報のうち、どのような視覚的特徴が各行動の決定に寄与しているかは定かでない。これを解明することで、人の視覚の解明だけではなく、コンピュータビジョンやロボットビジョンの分野、さらには人工網膜といった医療の分野にも貢献することができる。

ここで人工網膜とは、網膜色素変性症による失明者の網膜に埋植し電気刺激を与えることで、低解像度の映像を提示できる人工臓器のことである。人工網膜に関して、日本独自のシステムとしては「脈絡膜状経網膜刺激(Suprachoroidal-Transretinal Stimulation: 以下 STS)方式」が開発されてきた[2]。現在では 49(=7×7)極の電極を有しているため、失明患者に 49 画素の映像を提示することができる。

本稿では人工網膜への応用を想定して、リーチング動作に代表される目と手の協応運動に着目し、この動作で重要となる視覚的特徴を解明することを目的とする。その手法としては、被験者が装着しているビデオカメラの映像を低解像度の映像に変換して提示し、その視覚情報でリーチング動作を行ってもらおう。これにより、どのような視覚的特徴を取得し提示すれば、リーチング動作の精度が向上するかを確認できるため、リーチング動作で重要となる視覚的特徴を知ることができる。今回はその特徴抽出手法として、ヒトのボトムアップな視覚的注意の計算モデルであり、画像中で顕著な特徴を持つ場所を高い輝度値で表示する顕著性マップ[3,4]の使用が有効ではないかと考えた。すでに人工網膜への顕著性マップの応用は提案されている[5]が、異なる特徴抽出手法との比較は行われておらず、顕著性マップのどのような性質が優れているかは明らかでない。そこで、実験条件を人工網膜の映像提示条件と合わせ、輝度値のみを使用した変換処理と比較することで、人工網膜における顕著性マップの利点を確認し、リーチング動作に重要となる視覚的特徴を解明する。

2. 映像変換処理

被験者が身に付けたビデオカメラから取得される映像が、ヘッドマウントディスプレイ (Head Mounted Display: 以下 HMD) で提示される映像に変換されるまでの流れを図 1 に示す。ここで図 1 では、映像中の 1 フレームに対する処理を図示している。まず、ビデオカメラからは 640×480 画素の画像が得られる。そして、この画像に対して特徴抽出を行い、低

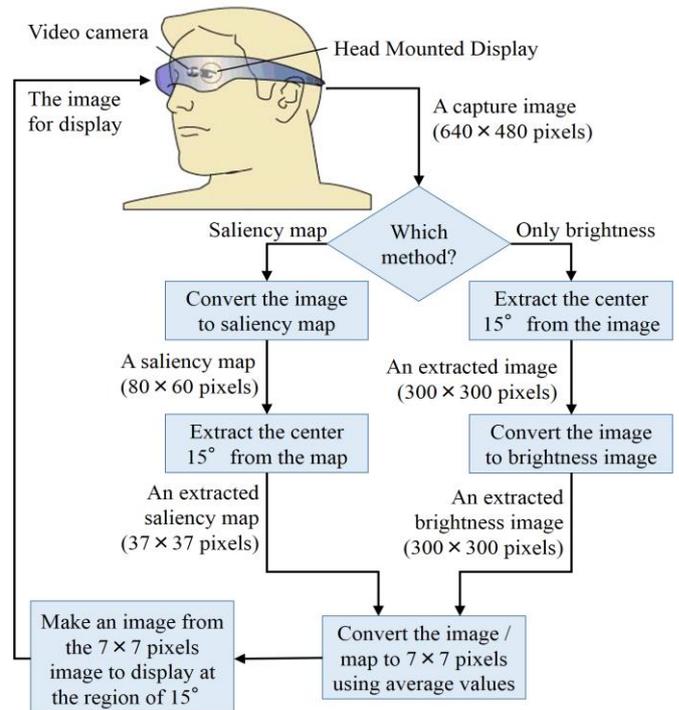


Fig.1 The process flow from an image recorded by the video camera to displaying the converted image to the head mounted display. After Fujikado (2006) with permission of Medical-Aoi Publications [6].

解像度の画像に変換する必要がある。ここで、その解像度を STS 方式で提示できる 7×7 画素とする。また、階調数は 8 とする。最後に、人工網膜により得られる視野角は $15 \sim 20^\circ$ 程度であるため、カメラ映像の 15° の範囲から変換されている 7×7 画素の画像を、HMD の約 15° の範囲に拡大し提示する。

特徴抽出の手法としては、顕著性マップを用いた手法と輝度値のみを用いた手法のどちらかを用いる。これらの手法の詳細を以下の節に記す。

2.1 顕著性マップを用いた手法

顕著性マップでは、輝度、色、線の方向、光の移動方向、光の点滅、これら 5 つの視覚的特徴を用いる[4]。輝度、色、線の方向は現在処理している画像のみから求めるが、移動方向と光の点滅は、現在の画像に加えて前フレームの画像が 1 枚必要となる。それぞれの特徴に対して、局所的に値の変化が大きい場所が強調されるが、その変化が大きい場所が画像全体に数多く存在するほど顕著性は減少していく。そして、これら 5 つの特徴に対して顕著性を求めた後、それらを線形結合することによって顕著性マップが得られる。詳細は文献[3,4]を参照されたい。

処理の手順としては、まず、ビデオカメラから取得された 640×480 画素の映像より、顕著性マップを求めることで 80×60 画素の映像を取得する。ここで、顕著性マップを求める際に画像の幅と高さが 8 分の 1 となっているが、この圧縮率はパラメータに依存する。そして、ここから事前調査により 15° の範囲であることがわかった中心の 37×37 画素を抽出し、これを 7×7 の領域に分割し、その平均値を計算する。最後にそれを 8 階調に変換することで、 7×7 画素の画像を作成する。

2.2 輝度値のみを用いた手法

顕著性マップの有効性を検証する比較手法として、輝度値のみの特徴を用いる。処理の手順としては、まず、ビデオカメラから取得された 640×480 画素の映像の中心より、事前調査により 15° の範囲であることがわかった 300×300 画素を切り出す。そして、この 300×300 画素の輝度値を取得し、これを 7×7 の領域に分割し、その平均値を計算する。最後にそれを 8 階調に変換することで、 7×7 画素の画像を作成する。

3. 実験内容

実験としては図 2 に示すように、 1280×1024 画素のタッチパネルディスプレイに 120×120 画素の正方形の視標をランダムな位置に表示させた。そして、被験者は 0.4m 離れた位置から、HMD に表示された 49 画素の情報を頼りに、その視標の中心に触れる。これにより、視標の中心と触れた位置のずれを測定した。そして、上記の試行 20 回を 1 セットとし、1 人の被験者に対して実験は 11 セット行った。ここで、この 11 セットを 3 つのパートに分類し説明していく。パート 1 はタッチパネルディスプレイの表示を白視標・黒背景（ディスプレイを黒色に表示し、その中に白色の視標を表示）もしくは黒視標・白背景として 2 種類、そして特徴抽出手法を顕著性マップと輝度値のみの 2 種類、これら組み合わせて 4 セットである。パート 2 はパート 1 の白と黒を輝度値の近い赤と緑に変更した条件であるため、同様に 4 セットとなる。パート 3 は特徴抽出を全く行わず、条件としては (1)HMD を着用しない、(2)HMD を着用して 15° の範囲でビデオカメラの映像を表示、(3)HMD を着用して何も表示しない、この 3 条件で 3 セットである。この実験は被験者 5 人に対して行った。

ここで、パート 1, 2 の実験目的は、環境が異なっても頑強

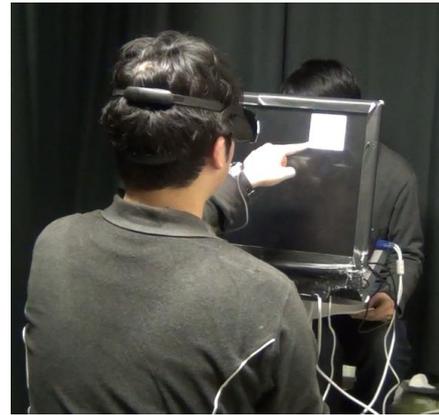


Fig.2 A scene at the time of an experiment.

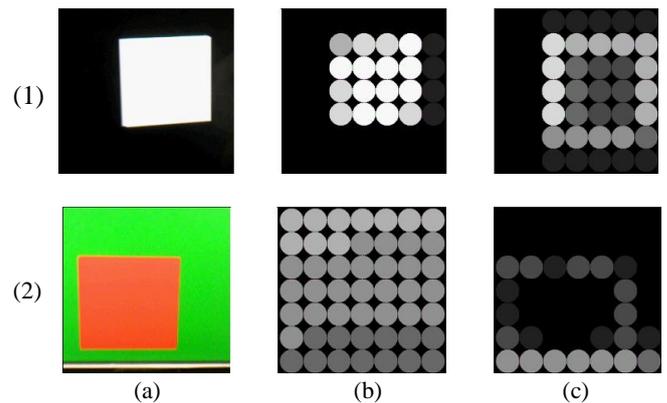


Fig.3 Appearance of a target. (a) Examples of a capture image. (b) Converted images using only brightness. (c) Converted images using saliency map. (1) White targets. (2) Red targets.

に視標や指を見分けられる視覚的特徴を得ることである。パート 1 の実験では、黒と白という明確に輝度値の異なる視標と背景であるため、輝度値のみでも十分精度良く中心に触れられ、逆に顕著性マップは輝度値以外の特徴も使用しているため精度が劣るのではないかと予想される。一方、パート 2 の実験では、輝度値の近い赤と緑という視標と背景であるため、輝度値のみでは視標の検出は困難であるのに対して、顕著性マップでは色や線の方向の特徴により、問題なくタスクを実行できることが期待される。また、パート 3 はそれ以前のパートの比較対象とするために実験を行った。

4. 実験結果および考察

輝度値のみ、もしくは顕著性マップを用いたときの視標の見え方の例を図 3 に示す。図 3(a)はカメラ映像の約 15° (300×300 画素) の範囲、図 3(b)は輝度値のみを用いて(a)を変換した映像、図 3(c)は顕著性マップを用いて(a)を変換した映像である。また、図 3(1)は白視標、図 3(2)は赤視標を用いた実験の映像であり、黒視標は白視標、緑視標は赤視標と同様の見え方であった。図 3 に示したように、白視標においては輝度値のみ、顕著性マップともに視標の位置が把握できた。一方で、赤視標においては図 3(2)-(b)に示したように、輝度値のみでは視標の位置がわからなくなるが、顕著性マップでは白視標と同様に視標の輪郭が表示されることがわかった。

パート 1, 2 の実験による測定結果を図 4 に示す。この結果に対して両側 t 検定を行い、 $p < 0.05$ を有意差ありとした。図 4 の結果から白視標のときに、予想に反して顕著性マップのほうがリーチングの精度が有意に良いことがわかった。また、

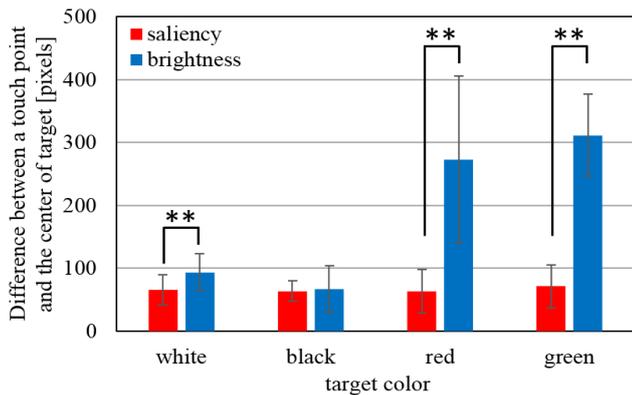


Fig.4 Results of a measurement and comparison between saliency and only brightness method (** paired t-test value $p < 0.05$).

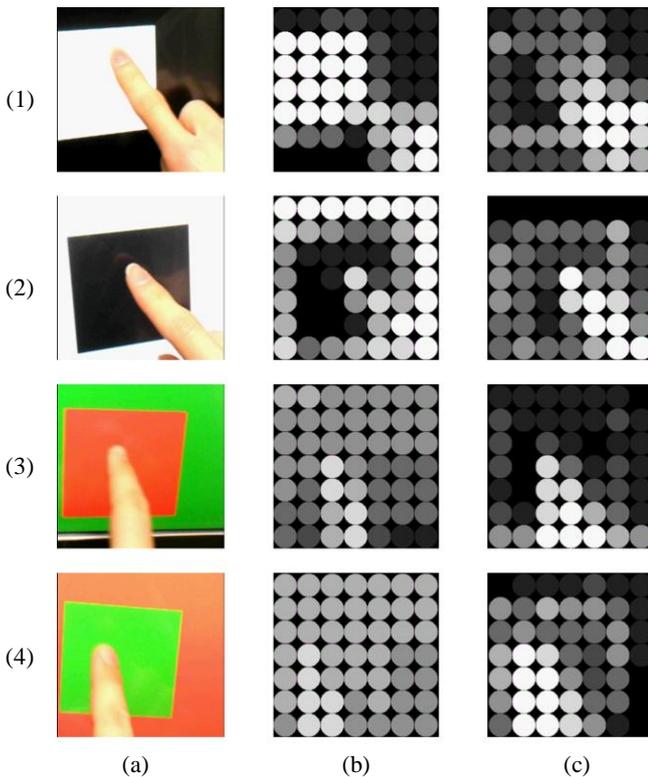


Fig.5 Appearance of a finger and a target. (a) Capture images. (b) Converted images using brightness. (c) Converted images using saliency map. (1) White, (2) black, (3) red and (4) green targets.

黒視標においても有意差はなく、この条件では一見無関係と思われる特徴を多く有している顕著性マップは、輝度値と比べて同等以上の性能を発揮することが確認できた。ここで、白視標のときに輝度値のみの精度が悪くなった原因としては、図 5(1)-(b)に示すように視標と指の色の輝度値の差が小さいため、視標を押そうとした指が視標と同化してしまい、指の位置がわからなくなりやすいからである。また、赤視標や緑視標を用いたときには予想通り、輝度値のみの視標の検出は困難となるのに対して、顕著性マップを用いたときには白と黒の視標の精度と遜色がない結果を得ることができた。一方で、パート 3 の結果である図 6 に示したように、HMD に何も表示させない条件では、視標の中心と触れた位置のずれの平均は約 474 画素であった。この結果と赤や緑視標の輝度値のみの結果を比較すると、後者のほうが有意にずれの少ない結果であった。この原因としては、カメラが移動することに

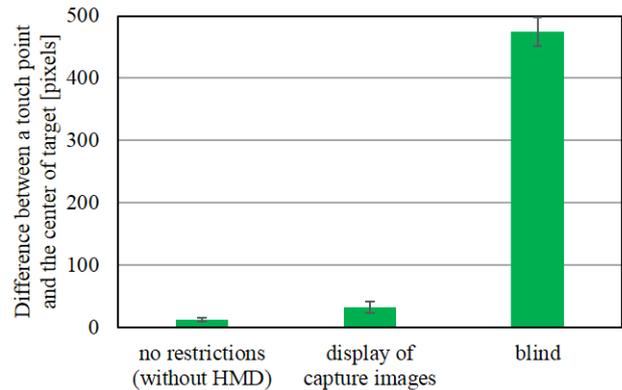


Fig.6 Results of a supplementary measurement.

よって照明環境が大きく変化したとき、輝度値の近い赤と緑であっても視標が見えることがあったためだと考えられる。

図 4 と図 5 を比較すると、視標と指のどちらの位置も把握しやすい場合にリーチングの精度が良いことがわかる。また、その精度は図 6 に示した HMD にカメラの映像を表示する条件と比べても、わずか 30~40 画素の差であるため、解像度の違いを考えると限界に近い結果が得られていると考えられる。

以上より、輝度値のみで達成できそうなタスクにおいても、それ以外の特徴を複合して使用していることがわかった。そして、白視標の条件において、指が見えなくなることによって輝度値のみの結果が有意に悪化したため、その他の特徴が指の位置を把握するために必要な情報を含んでいるといえる。また、顕著性マップを用いた場合には、環境が変化しても視標や手は同様に見えることより、環境の変化にも頑強であることがわかった。したがって、複数の特徴を統合でき、環境の変化にも頑強である顕著性マップは、目と手の協応運動において重要な視覚的特徴を与えるモデルであるといえる。

5. 結論と今後の課題

今回の実験より、顕著性マップは環境の変化に対して頑強に視標や指の位置が把握できるモデルであることがわかった。また、輝度値のみを用いた手法よりも顕著性マップを用いた手法のほうがリーチングの精度が良かったため、人工網膜への応用が期待される。今後の課題としては、顕著性マップの 5 つの特徴のうち、どの特徴が指の位置の把握に寄与しているかを調査していく必要がある。また、今回は被験者が 5 名のみだったため、被験者数を増やしていく必要もある。さらに、今回は特にリーチング動作について実験を行ったが、歩行といった他の動作に必要な視覚的特徴も解明していきたい。

文献

- [1] 教育機器編集委員会編, "産業における教育訓練と教育機器", 産業教育機器システム便覧, 日科技連出版社, pp.4-31, 1972.
- [2] 神田寛行, 不二門尚, "感覚系における人工臓器—人工網膜—", 人工臓器, vol.41-3, pp.202-206, 2012.
- [3] Itti, L., Koch, C. and Niebur, E., "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.20-11, pp.1254-1259, 1998.
- [4] Itti, L., Koch, C. and Niebur, E., "Realistic avatar eye and head animation using a neurobiological model of visual attention", in *Proc. SPIE 48th Annu. Int. Symp. Opt. Sci. Technol.*, pp.64-78, 2003.
- [5] Parikh, N. et al., "Performance of visually guided tasks using simulated prosthetic vision and saliency-based cues", *Journal of Neural Engineering*, vol.10-2, pp.1-13, 2013.
- [6] 不二門尚, "網脈絡膜変性疾患の治療に向けて: 人工網膜", あたらしい眼科, vol.23-9, p.1170 (図 1), 2006.