

## 動画像コンテンツにおける注視点マップと顕著性マップ との関係性に関する考察

### Consideration on the Relationship between Gaze Map and Saliency Map on a Natural Scene

上向 俊晃<sup>†</sup>      小峯 一晃<sup>‡</sup>      森田 寿哉<sup>‡</sup>  
Toshiaki Uemukai<sup>†</sup>    Kazuteru Komine<sup>‡</sup>    Toshiya Morita<sup>‡</sup>

#### 1. はじめに

画像に対する注意の向けられやすさ（顕著性）の分布を算出する「顕著性マップ」が提案されている。これは、画像データに含まれる物理的特徴を解析するだけで求めることができる。しかし、実際に測定した視線では、人間が注意を向けている位置と視線が向けられている位置はほぼ一致するが、顕著性マップにおいて顕著性の高い位置に、必ずしも注意が向けられるとは限らない。その原因の一つとして、複数の物理的特徴の解析結果を合成して顕著性マップを算出する際のパラメータの設定方法に関して、十分に検討されていないことが挙げられる。

そこで、本研究では、パラメータ最適値を求めることを目的として、動画像コンテンツを視聴する被験者の視線が向けられる画像上の位置（注視点）を元に算出した注視点マップと、注意の向けられやすい領域を示すとされる顕著性マップとの関係性について考察する。具体的には、視線計測実験結果から得られた注視点マップと顕著性マップとの一致度を指標として、顕著性マップにおける重みパラメータの影響を調べる。

#### 2. 背景

##### 2.1 動画像コンテンツに対する評価

動画像コンテンツ制作者（ディレクターやカメラマン）は、視聴者に制作意図が伝わりやすくするために、視聴者の注意が向けられるような演出上の工夫をしている。また、動画像コンテンツ提供者も、挿入する広告などに対する視聴者の注意に関する関心は高い。ここで、制作者は、過去の経験や専門技法に基づいて、構図やカメラワークなどを決定することが一般的である。しかし、制作した動画像コンテンツが、視聴者にどのように見られているのか、あるいは、注意が向くようにするには、どのような構成にすればよいかなどを知ることができれば、有効である。動画像コンテンツを視聴者に実際に見せて、その視線の動きや瞬目から注意の程度を評価する技術[1]も存在するが、装置が大規模であり、労力と手間がかかる。被験者を使わず、制作した動画像コンテンツ自体を解析するだけで、注意の向けられやすさを評価することが効率性の点で望ましい。

##### 2.2 顕著性マップ

画像に対する注意の向けられやすさ（顕著性）の分布を算出する「顕著性マップ」[2][3]が提案されている。これは、被験者を必要とせず、画像の物理的な特徴を解析

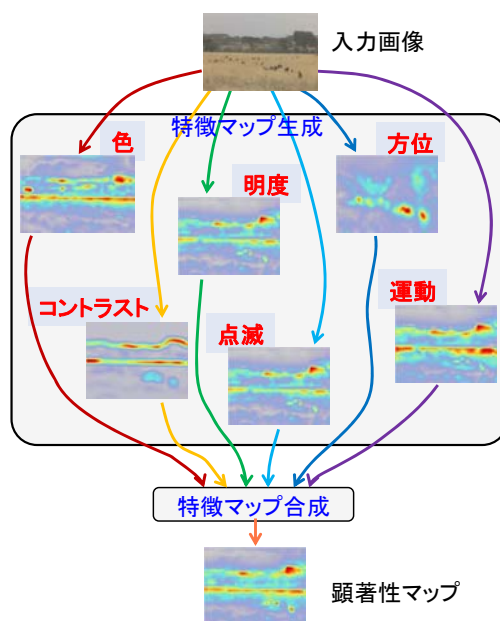


図1: 顕著性マップの概念図

するだけで注意が向けられる位置を推定することができる。図1は、顕著性マップの算出過程に関する概念図を示す。顕著性マップの算出過程において、「特徴マップ生成」と「特徴マップ合成」がある。「特徴マップ生成」では、画像データに対して、6つの視覚属性それぞれに関する画像解析を行う。視覚属性には、色(Color)、明度(Intensity)、方位(Orientation)、コントラスト(Contrast)、点滅(Flicker)、運動(Motion)がある。さらに、「特徴マップ合成」では、視覚属性ごとに生成した特徴マップを線形和し、顕著性マップを算出する。このように、顕著性マップは、画像の物理的な特徴から注意の向けられやすい領域を求めることができるが、実際にその画像を人間に見せたとき、顕著性の高い領域に、人間の注意が向けられるとは限らない。

また、顕著性マップの算出過程において、各特徴マップを線形和にて合成する際、各視覚属性に重みパラメータを付与することができる。すなわち、同じ動画像データであっても、これらの重みパラメータを変化させることで、最終的な顕著性が変わることになる。注意の特性に応じた最適な重みの配分があると考えられるが、これらについては十分な検討がなされていない。

そこで、本稿では、動画像コンテンツに対する注視点の分布と、動画像データの物理的な特徴のみから算出される顕著性マップとの一致度を指標として、顕著性マップにおける重みパラメータの影響を調べ、重みパラメータの最適化手法を検討する。

<sup>†</sup> (株) KDDI 研究所, KDDI R&D Laboratories, Inc.

<sup>‡</sup> NHK 放送技術研究所, NHK Science & Technical Research Laboratories

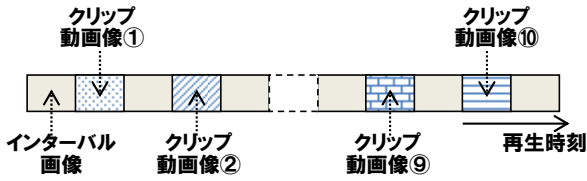


図 2: 1つの試行動画像の構成

### 3. 動画コンテンツを利用した視線計測実験

本章では、パラメータ最適化手法の検討のために行った視線計測実験とその結果について説明する。

#### 3.1 実験方法

視聴者が注意を向けている領域は、多くの場合、視線が向けられている領域であると言える。しかし、顕著性マップにおいて顕著性の高い領域に、注意が向けられるとは限らない。これは、人間が注意を向けるとき、コンテンツ自体の視覚的な特徴（ボトムアップ要因）だけでなく、視聴者の嗜好、興味、経験、知識、文脈理解（トップダウン要因）にも影響されるためである[4]。例えば、本研究の先行研究[1]で行った視線計測実験では、2人のキャストが映っている動画像において、注視点は、キャストの顔や口元にあることが分かった。しかし、同じ動画像データから算出した顕著性マップでは、キャストの頭部や、キャストの背後に設置された番組セットの一部における顕著性が高くなり、注視点とずれることが分かった。

そこで、本研究では、トップダウン要因を可能な限り排除できるように、短時間の動画コンテンツを提示し、そのときの視線を計測した。具体的には、実際に放送された約 20 番組から、5 秒音声なしクリップ動画像（30fps）を 100 種類抽出した。ここで、5 秒と短く設定したのは、被験者に考える時間を与えず、画像から自然に誘導される視線の動きを捉えるためである[5]。さらに、図 2 に示すように、クリップ動画像 10 個を 1 組とし、5 秒のインターバル画像を、各クリップ動画像の前後に挟み、合計で 1 分 45 秒の試行動画像を作成した。このように作成した 10 個の試行動画像を各被験者に提示した。各クリップ動画像の画像フレームごとに、注視点の座標を求め、合計 80 名の被験者に対して実験を行った。

実験では、瞳孔角膜反射法で視線計測を行った。ディスプレイ（サイズ:100inch、解像度:1920x1080）から所定距離（約 3m）だけ離れた座席に着席（画角:水平約 41 度、垂直約 24 度）した視聴者の眼球部分を、近赤外線光で照射し、それをカメラで撮影する。撮影された瞳孔および角膜反射像から視線の向きを計測し、動画像の表示領域内の注視点を求める。

#### 3.2 実験結果

図 3(2)は、あるクリップ動画像の 1 枚の画像フレーム（図 3(1)）に対して、計測できた注視点の位置に赤丸を重畳してプロットした例を示す。なお、計測した注視点の座標データには、瞬目などの影響で、画像フレームの表示領域の外に外れた点は除去した。



(1) 原画像 (2) 注視点のプロット

図 3: 注視点のプロット例

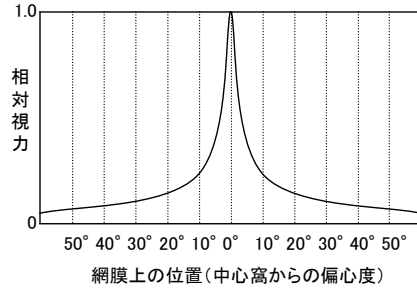


図 4: 網膜偏心度と視力の関係

### 4. 注視点マップと顕著性マップとの比較手法

本章では、3 章で計測した注視点の分布「注視点マップ」と、動画コンテンツの物理的特徴から算出した顕著性の分布「顕著性マップ」とを、双方の分布の重なり度合い（一致度）を指標として比較する。

#### 4.1 注視点マップ

注視点マップと比較する顕著性マップは、画像フレーム上の全画素に対して、顕著性の程度を数値化したものである。一方、視線計測実験によって得られる注視点は、画像フレーム上の 1 点（1 画素）であるため、画像フレーム上の全画素に対する注視点の集中度を数値化するには、膨大な数の被験者を集めた視線計測実験を行う必要があり、非現実的である。そこで、顕著性マップと比較しやすく、かつ、少数の被験者から得られた注視点からでも、画像フレーム上の全画素に対する注視点の集中度を数値化するため、「視力分布」の概念を導入して注視点マップを作成する。

$$V = \frac{Vf}{1 + \frac{E}{Es}}$$

文献[6]によると、「視力分布」の概念として、眼球の網膜偏心度と視力の間には、図 4 のような関係があり、以下の式が成立する。Vfは中心窩での視力、Esは定数とすると、Vは偏心度Eでの視力を示す。

これは、図 5 に示すように、視線の向きを中心としたときの視力に対して、ある偏心度 E の方向での視力の相対値を示す。視線の向きからの偏心度が大きくなるにつれて、視力が指数関数的に低下することを示している。

本研究では、視線計測実験において計測された注視点の位置が中心窩であると想定し、注視点における視線の集中度を Vf、注視点(GX,GY)から距離 R の位置にある任意の点(X,Y)における視線の集中度を V として、上記式に当てはめることで、注視点マップを作成する。なお、偏心度 E に関しては、被験者の眼球位置注視点を結ぶ視線の向きは、常に動画像表示ディスプレイと垂直の関係にあると近似し、眼球位置と表示ディスプレイ間の距離 L

$$V[X, Y] = \frac{Vf}{1 + \frac{E[X, Y]}{Es}} \quad 1 \leq X \leq W \quad 1 \leq Y \leq H$$

$$E[X, Y] = \text{atan} \frac{R[X, Y]}{L}$$

$$R[X, Y] = \sqrt{(X - GX)^2 + (Y - GY)^2}$$

と注視点と点(X,Y)との距離 R から求めた。つまり、画像フレームの解像度を幅 W 高さ H とすると、以下の式で求められる。

各画像フレームに対して、被験者ごとの注視点マップを求める。さらに、計測できた全被験者に関して注視点マップを線形和することで、各画像フレームに対する注視点マップを作成する。図 6 に例を示す。図では、注視点の集中度が高い領域は赤色、低い領域は青色で示す。

上記の結果、注視点マップを、行数 H 列数 W の行列データとして算出する。つまり、ある画像フレーム f において、被験者 s の注視点マップ GMs(f) とすると、全被験者分の注視点マップ GM(f) を、以下のように求める。

$$GM(f) = \sum GMs(f)$$

$$GMs(f) = \begin{bmatrix} V[1,1] & V[2,1] & \dots & V[W,1] \\ V[1,2] & V[2,2] & \dots & V[W,2] \\ \vdots & \vdots & \ddots & \vdots \\ V[1,H] & V[2,H] & \dots & V[W,H] \end{bmatrix}$$

#### 4.2 顕著性マップ

動画像を構成する各画像フレームに対する顕著性マップを求める。このとき、各視覚属性に対する特徴マップの合成において付与可能な重みパラメータを設定する。なお、顕著性マップの算出には、「Graph-Based Visual Saliency」[7][8]を利用する。

上記の結果、顕著性マップを行数 H 列数 W の行列データとして算出する。つまり、ある画像フレーム f において、顕著性マップ SM(f) を以下のように求める。

$$SM(f) = \begin{bmatrix} F[1,1] & F[2,1] & \dots & F[W,1] \\ F[1,2] & F[2,2] & \dots & F[W,2] \\ \vdots & \vdots & \ddots & \vdots \\ F[1,H] & F[2,H] & \dots & F[W,H] \end{bmatrix}$$

#### 4.3 一致度

注視点マップと顕著性マップ間的一致度を次のように定義する。注視点マップ行列 GM と顕著性マップ行列 SM において、i 列 j 行 (1 ≤ i ≤ W, 1 ≤ j ≤ H) の要素同士の差 CM の絶対値が閾値 Th よりも小さい [i,j] の組数 count を求める。その結果、一致度 F は、以下のように求める。つまり、一致度は、全要素数に対して、差分値 CM が

$$CM[i, j] = |SM[i, j] - GM[i, j]|$$

$$\begin{cases} count = 0 & (\text{初期値}) \\ count = count + 1 & (CM[i, j] \leq Th \text{ の場合}) \\ count = count & (CM[i, j] > Th \text{ の場合}) \end{cases}$$

$$F = \frac{count}{W \times H}$$

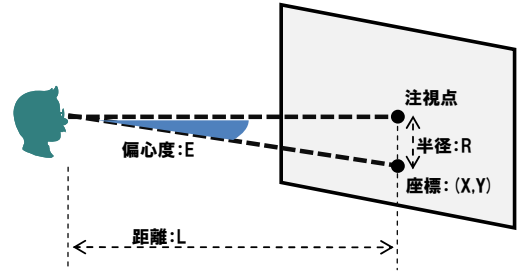


図 5: 注視点と偏心度

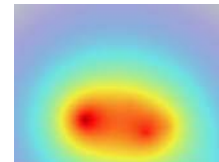


図 6: 図 3 に対する注視点マップの例

閾値 Th よりも小さい要素数の割合を示し、F 値が大きいほど、注視点マップと顕著性マップの重なり度合いは高いことを示す。

#### 4.4 比較結果

表 1 に示す内容を持つ 4 つの動画像コンテンツに対して、画像フレームごとの注視点マップ GM と顕著性マップ SM から一致度 F を求めた。このとき、表 2 に示す重みパラメータの設定パターンにて、顕著性マップを算出した。4.3 節における一致度を求める際、GM と SM の行列サイズは H=240, W=320 とし、それぞれの全要素において、最大値を 1、最小値を 0 となるように正規化した上で、差分値 CM を求めた。また、CM の全要素において、最大値を 1、最小値を 0 となるように正規化した上で、閾値 Th を 0.6 に設定し、F を求めた。さらに、全画像フレーム (150 枚) における F の平均値 F' を求めた。

本コンテンツを利用した 80 名の被験者による視線計測実験の結果、動画像コンテンツを構成する 150 の画像フレームそれぞれに対して、コンテンツ A の場合は、最大 36 名分、最小 16 名分の注視点、コンテンツ B の場合は、最大 41 名分、最小 14 名分、コンテンツ C の場合は、最大 34 名分、最小 17 名分の注視点、コンテンツ D の場合は、最大 34 名分、最小 19 名分の注視点を測定できた。

図 7 に、コンテンツ D において、6 つの原画像に対する注視点のプロット、注視点マップ、顕著性マップの出力結果例を示す。図において、注視点のプロットは、当該画像フレームに対して測定できた全被験者の注視点の位置にプロットしたものの、注視点マップは、各被験者の注視点に 4.1 節で述べた視力分布の概念を適用したマップを全被験者分総和したものを示す。

図 8 に、コンテンツ A からコンテンツ D それぞれで、視覚属性の重みパラメータのパターンを変化させたときの一貫性を示す。各グラフの縦軸は F' 値を示し、横軸は表 2 に示した重みパラメータの設定パターンを示す。

また、図 9 に、コンテンツ A とコンテンツ B のそれぞれに対して、各視覚属性の重みパラメータの設定値を変化させたときの一貫性の変化について求めた結果を示す。図中の「C」は、「色」視覚属性の重みパラメータだけを 0.1、0.5、1.0、2.0、5.0、10.0 と変化させ、他の視覚属性の重みパラメータを 1.0 としたときの一貫性を示す。同

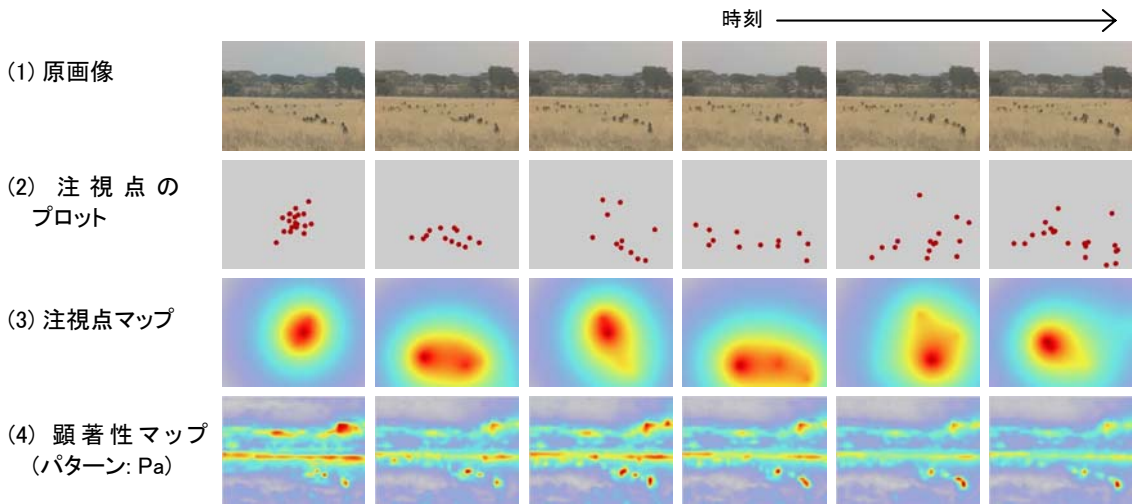


図7:コンテンツ D に対する結果例

表 1: 動画コンテンツの内容

種別	内容
A	CGアニメーションであり、シリンダーの上下運動に伴う煙の動きや、カメラズームが含まれる。
B	中学生が携帯電話のカメラでもう1人の中学生を撮影している映像であり、カメラパンが含まれる。
C	教室の中を小学生がカメラの方向に向かってくる映像であり、カメラズームが含まれる。
D	草原の中を動物がゆっくり歩く様子を撮影した映像であり、カメラの動きはまったくない。

表 2: 重みパラメータの設定パターン

種別	内容
Pa	6つの視覚属性の重みパラメータを1.0に設定
PeC	「色」属性のみ1.0、他の属性を0に設定
PeI	「明度」属性のみ1.0、他の属性を0に設定
PeO	「方位」属性のみ1.0、他の属性を0に設定
PeR	「コントラスト」属性のみ1.0、他の属性を0に設定
PeF	「点滅」属性のみ1.0、他の属性を0に設定
PeM	「運動」属性のみ1.0、他の属性を0に設定

様に、「I」は明度、「O」は方位、「R」はコントラスト、「F」は点滅、「M」は運動の一致度を示す。各グラフの縦軸は F 値を示し、横軸は各視覚属性の重みパラメータの設定値を示す。

図 8、図 9 から、視覚属性の重みの設定パターンに応じて、一致度の値が変化している。その中で、方位の視覚属性(PeO)の一致度が最も高いことが分かる。コンテンツ B や C のように、建物内で人物を撮影し、さらにカメラ動作が含まれる映像の場合、各画像フレームは細かく複雑な線画を多く含むため、方位の視覚属性による顕著性が高い領域が広範囲に存在した。また、コンテンツ D のように、草原を遠方から定点カメラで撮影した映像の場合、各画像フレームは線画も少なく、色の変化も少ない中でも、被験者の視線を集めやすい動物の領域に対して、顕著性が高くなった。一方、図から、運動の視覚属性(PeM)の一致度が最も低いことが分かる。コンテンツ A、B、C のように、カメラ動作が含まれたり、被写体が動く映像よりも、コンテンツ D のように、カメラ動作がなかったり、被写体の動きがあまりない映像の方が、一致度は低くなった。被写体に動きのある映像の場合、それに視線が追従することによって、一致度が高くなることが予想されるが、本評価では、全画像フレームに対する一致度の平均値を採用しており、一瞬の動き(部分的な画像フレーム)に対する一致度で行っていないこと、また、本評価では、画像フレームの表示タイミングと視線データの記録タイミングが同じであると想定し、フレーム画像提示から注視までの時間の遅れを想定しなかったため

と考えられる。これらの影響についての評価は、今後の課題である。

## 5. 考察

上記の結果を踏まえて、考察する。

### 5.1 比較結果について

4.4 節の比較結果から、異なる動画コンテンツにも関わらず、注視点マップと顕著性マップとの一致度の特性が類似していることから、トップダウン要因による注視点に対する影響が、抑制できている可能性が高いと考えられる。また、各視覚属性の重みパラメータの中で、方位の視覚属性の一致度が最も高いことが分かった。

本研究では、1つの視覚属性単独での影響について評価した。そこで、今後は、すべての視覚属性の影響を総合的に考慮し、注視点マップと顕著性マップとの一致度を最も高くする最適な重みパラメータ値の算出方法について検討する予定である。つまり、6つの視覚属性に対して、6次元空間でのパラメータ最適解を算出することを考えている。

### 5.2 注視点マップ作成方法について

本研究では、注視点マップは、画像フレームごとに注視点群の集中度合を視力分布という空間的な概念を利用して求めた。しかし、時間経過に伴う注視点の動き(スキャンパス)は、被験者ごとに異なり、同じ位置に停留したり、大きく変位したりを繰り返す。

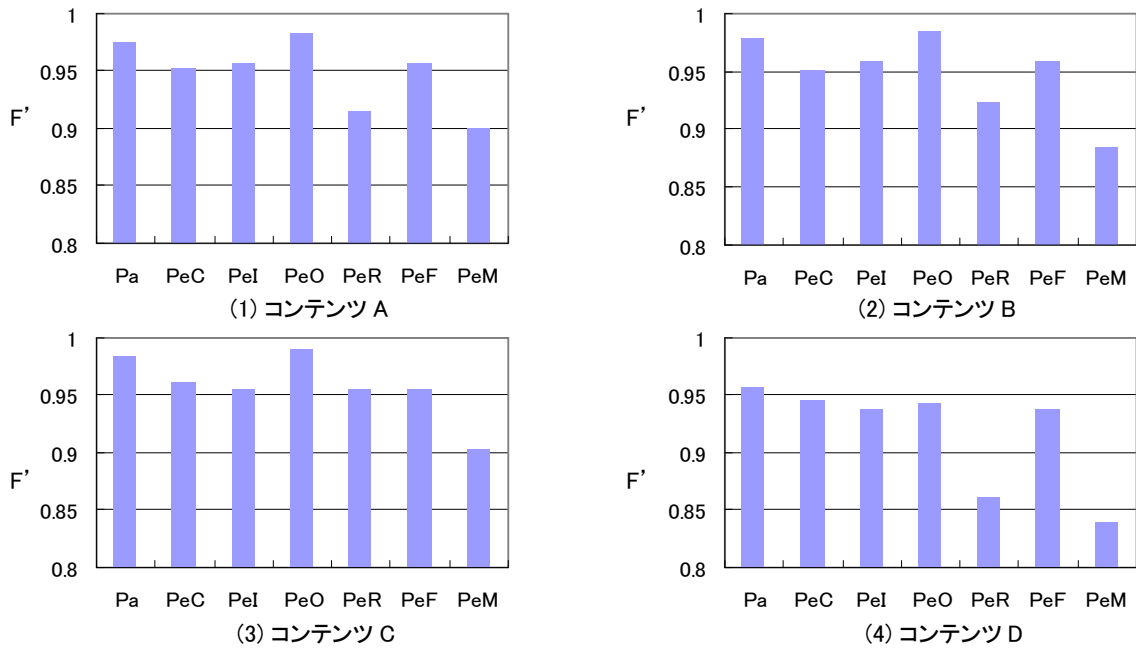


図 8: 各視覚属性の重みパラメータセットに対する一貫性

つまり、ある時点で同じ位置に注視点が集まっていますが、被験者ごとに注視した順番は異なる。順番の早い注視点の方が、優先度が高いと考えることができる。そこで、今後は、ある時点での注視点マップを生成する際には、各注視点のキャンパス上の順番と時間を考慮し、前後のつながりがある注視点マップを作成することを考えている。

### 5.3 一貫性算出方法について

本研究では、各注視点マップに含まれる注視点データは、比較対象となる顕著性マップの元となる画像フレームの表示時刻に計測したものを利用した。ところが、被験者の視線は、画像を見たのち、その画像の物理的特徴に刺激されて動くと考えられる。つまり、ある画像フレームに対する顕著性マップと比較すべき注視点データは、時間的に後方で計測されたものを利用することが有効であると考えられる。そこで、今後は、画像フレーム提示から注視までの時間の遅れを考慮した一貫性の算出方法について検討する必要がある。

また、本研究では、各画像フレームに対応する注視点マップと顕著性マップとの一貫性 $F$ を求め、動画コンテンツ再生区間の全画像フレームに対応する一貫性の平均値 $F'$ を評価値とした。図 10 は、コンテンツAとコンテンツBのそれぞれにおいて、画像フレームごとの一貫性の変化を示す。縦軸は画像フレームに対する一貫性 $F$ 、横軸はフレーム番号を示す。なお、顕著性マップ算出時の重みパラメータの設定は、設定パターンPaを利用した。図からも分かるように、画像フレームの進行に対する一貫性の推移は大きい。これは、サッカード途中も注視点としてカウントされることも原因の一つとして考えられる。したがって、視覚属性の重みパラメータの最適解を求めるには、動画コンテンツ再生区間全体での一貫性ではなく、一部区間の一貫性の推移に限定する必要があると考えられる。特に、サッカードは、視覚的な刺激を受け

て誘発されるため、サッカードが発生する直前の画像フレームと、サッカードで遷移した注視点との関係に注目することが有効であると考えられる。文献[9]でも、連続サッカードに注目し、時間経過と共に、ボトムアップ要因による影響が下がることも述べられている。

さらに、本研究では、注視点マップと顕著性マップとの関係を検証するため、差に基づく一貫性の指標を利用したが、DOH[9]やROC[10]などの指標を利用した既存研究も存在しており、顕著性マップにおける重みパラメータ最適値算出にふさわしい指標についても今後検証すべきである。

### 5.4 顕著性マップについて

顕著性マップは、画像データに含まれる色や明度などの物理的特徴のみに基づいて、画像データに対する注意の強度の分布を算出したものである。一方、本研究で行った視線計測実験では、5秒の動画を連結した動画コンテンツを利用することで、トップダウン要因による影響を抑制した。しかし、5秒であっても、提示された動画に対して、人間は映し出された動画の内容について瞬時に判断し、「被写体は人物である」、「被写体は花である」、「人物の性別は男性である」、「人物の顔は笑っている」などの識別がある程度可能である。その判断が可能かどうかは、被験者ごとに異なり、これまでの経験や知識、判断能力などに影響されると思われる。本研究で行った視線計測実験においても、動画中に現われた人物の顔や、挿入されたテロップなどの文字に対して、視線が集中していたことを確認している。

このように、注視点マップは、少なからずトップダウン要因に影響されてしまい、完全に顕著性マップと一致することはない可能性が残る。ただし、近年の画像処理、画像認識技術の発展により、画像内のオブジェクトが何であるかを認識することも可能である。特に、人物の顔の検出・認識技術は高精度であり、一般的なオブジェク

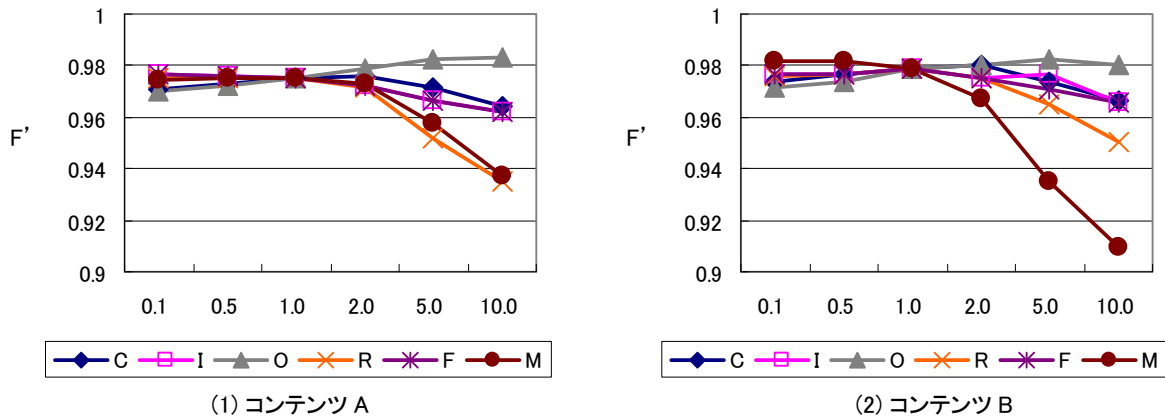


図 9: 視覚属性ごとの一致度の変化

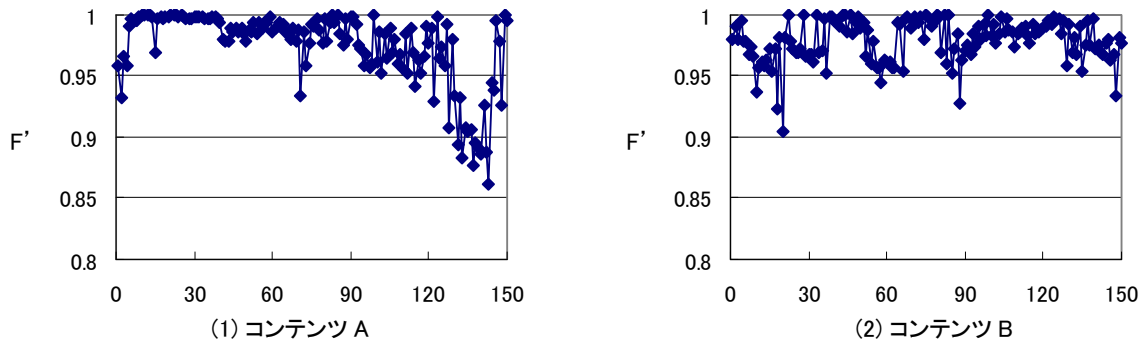


図 10: フレーム画像ごとの一致度の変化

トであっても、ある程度の検出・認識が可能である。そこで、顕著性マップの算出において、視覚属性の重みパラメータの最適値によって計算された顕著性マップを、画像フレームに対する画像認識結果に基づいて補正することなどが効果的であると考えられる。文献[11]では、顔検出・認識結果を考慮した顕著性マップを生成することで、注視点の位置予測を行っているが、重みパラメータの設定方法までは検討していない。

## 6. まとめ

本研究では、顕著性マップ算出時の視覚属性パラメータの設定に関して、視線計測実験結果から得られた注視点マップと顕著性マップとの一致性を指標にした検討を行った。その結果、視覚属性の種別に応じて、一致度が変化することを確認することができ、視覚属性ごとの一致度では種々のコンテンツに対して同様の傾向が得られた。また、一致度の算出方法や、注視点マップおよび顕著性マップの算出方法について考察した。

今後は、時間軸を考慮した注視点マップ作成方法、画像認識結果を考慮した顕著性マップ作成方法、画像フレーム提示から注視までの時間の遅れを考慮した一致度算出方法を含め、視覚属性重みパラメータの最適解を算出する方法に関して検討を行う予定である。

### 参考文献

[1] 澤島 康仁, 小峯 一晃, 比留間 伸行, 伊藤 崇之, 渡辺 誓司, 鈴木 祐司, 原 由美子, 一色 伸夫, “番組視聴時の視線分布と番組内容理解度の関係”, 映像情報メディア学会誌, Vol. 62, No. 4, pp. 587-594 (Apr. 2008).

[2] L. Itti, C. Koch, and E. Niebur, “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis”, Proc. of IEEE Transactions of Pattern Analysis and Machine Intelligence, Vol. 20, No. 11, pp. 1254-1259 (Nov. 1998).  
 [3] 小池 耕彦, 伊丸岡 俊秀, 齋木 潤, “顕著性マップ”, 心理学評論, Vol. 46, No. 3, pp. 391-411 (Jan. 2003).  
 [4] J. M. Henderson, “Human Gaze Control during Real-world Scene Perception”, TRENDS in Cognitive Sciences, Vol. 7, No. 11, pp. 498-504 (Nov. 2003).  
 [5] 山田 光穂, 福田 忠彦, “画像における注視点の定義と画像分析への応用”, 電子情報通信学会論文誌, Vol. J69-D, No. 9, pp. 1335-1342 (Sept. 1986).  
 [6] J.M. フィンドレイ, I.D. ギルクリスト, “アクティブ・ビジョン - 眼球運動の心理・神経科学”, 北大路書房 (Sept. 2006).  
 [7] J. Harel, C. Koch, and P. Perona. “Graph-Based Visual Saliency”, Proc. of Neural Information Processing Systems (NIPS), pp. 545-552 (Dec. 2006).  
 [8] “Saliency Map Algorithm: MATLAB Source Code”, <http://www.klab.caltech.edu/~harel/share/gbvs.php>.  
 [9] R. Carmi, L. Itti, “Visual Causes Versus Correlates of Attentional Selection in Dynamic Scenes”, Vision Research, Vol. 46, No. 26, pp. 4333-4345 (Dec. 2006).  
 [10] B.W. Tatler, R.J. Baddeley, I.D. Gilchrist, “Visual Correlates of Fixation Selection: Effects of Scale and Time”, Vision Research, Vol. 45, No. 5, pp. 643-659 (Mar 2005).  
 [11] M. Cerf, J. Harel, W. Einhaeuser, and C. Koch, “Predicting Human Gaze using Low-level Saliency Combined with Face Detection”, Proc. of Neural Information Processing Systems (NIPS), pp. 241-248 (Dec. 2007).