

注視行動の予測に向けた
 動画コンテンツの視覚的特徴と実測した注視点の正準相関分析
 Canonical Correlation Analysis between Visual Features of Videos and Human Gaze
 toward Prediction of Human Attention

上向 俊晃[†] 加藤 恒夫[†] 梅田 修一[‡] 小峯 一晃[‡] 森田 寿哉[‡]
 Toshiaki Uemukai[†] Tsuneo Kato[†] Shuichi Umeda[‡] Kazuteru Komine[‡] Toshiya Morita[‡]

1. はじめに

テレビ番組や広告映像などの動画コンテンツ提供者(制作者や販売者)にとって、撮影時の演出や映像制作時の編集が及ぼす、視聴者の注意の変化は重要な関心事である。コンテンツ提供者が映像内に込めた意図が伝わりやすいように注意を喚起できれば、番組の分かりやすさや広告効果など、コンテンツ品質の向上が期待できる。あるいは、視聴者の注意を引かず、期待した効果が得られなかった映像を客観的に分析できれば、次の演出や編集に活かせる。

人間の注意の方向は、視線と密接な関係がある[1]。実際に複数の視聴者の視線の動きを計測して、注視点の分布と視聴者の動画コンテンツの内容理解度との関係性を評価した例がある[2]。しかし、視線計測技術では、コンテンツを制作するたびに被験者を募り実験する必要があるため、労力と手間がかかるという課題があり、より簡単な評価方法が望まれる。

一方、画像データの物理的特徴量から注意の向けられやすい領域を推定する「顕著性マップ」[3][4]が注目されている。これは、1つの画像データに対して、複数の物理的特徴で解析した特徴マップを生成した上で、それぞれの特徴マップの重み付き線形和として算出できるため人間の注意モデルに基づいた画像データの評価を行える。

動画コンテンツを利用して、顕著性マップで得られた顕著性の高い領域に、視線計測実験で得られた注視点があれば、動画コンテンツの視覚的特徴と、視聴者の注視行動は相関が高いと言える。しかし、実際に視聴者の視線計測を行った結果、注視点の分布は、顕著性の高い領域と必ずしも一致しないことが分かっている。これは、被験者の注意は、画像の物理的特徴(ボトムアップ要因)だけでなく、嗜好、経験、知識などに基づく瞬間的な文脈理解(トップダウン要因)に影響されること[5][6]が主な原因である。そこで、トップダウン要因を排した条件で視線計測実験を行った[7][8]。しかし、それでも注視点の分布と顕著性の高い領域とは一致しない。従来研究では、顕著性マップ算出時に用いる重みパラメータを、暫定的に初期値のまま設定した場合や、少ない実験例から経験的に設定した場合[9]が多く、その最適値を体系的に求める手法はまだ確立されていない[5]ことも原因である。[9]では、動画コンテンツにおける視聴者の注意を引く特徴を定量化することで、動画コンテンツの中で注視されるべき領域を推定する手法が提案されている。実際に動画コンテンツを視聴した視聴者の眼球運動を計測して求めた注視領域と、動画コンテンツの物理的特徴から推定した注視領域が一致するように、擬似逆行列を用いて重みパラメータを求めている。

しかし、重みパラメータは、動画コンテンツに依存していると結論付けられているが、ある重みパラメータが、どのような内容あるいは物理的特徴を持つ動画コンテンツに適しているかなど、より深い議論までには至っていない。

本稿は、トップダウン要因を排した条件下において、特徴マップの重みパラメータ最適化により、どれだけ視線を推定できるのかを評価することを目的とし、動画コンテンツから得た顕著性マップと、視線計測実験から得た注視点の分布との相関関係を分析する。このとき、顕著性マップが特徴マップの線形和で得られることに着目して、正準相関分析を用いる。これにより、動画コンテンツの視覚的特徴と視聴者の注視行動との相関関係が高くなるような重みパラメータを導出する。

まず、提案手法では、トップダウン要因をできるだけ排除するために、短時間の動画コンテンツを提示する視線計測実験で得られた注視点(実注視点)を利用する。また、顕著性マップは、解析対象画像データの解像度分の膨大な情報量を持ち、実注視点の数(情報量)と大きく異なる。そこで、提案手法では、顕著性マップが持つ情報から、視線の推定に役立たない情報を削除し、役立つ情報のみを抽出するため、人間の注意が向けられやすい位置として複数の顕著性極大点を求める。さらに、顕著性極大点における視覚的特徴を示す「視覚的特徴ベクトル」を求める。これは、顕著性極大点における物理的特徴量を各要素に持つ。また、顕著性極大点周辺にどれだけ実際の視線(実注視点)が集まったかを示す「注視点集中度」を求める。これは、実注視点と顕著性極大点との、視線計測実験誤差も含めた一致の度合いを示す。顕著性極大点における顕著性が高いほど、顕著性マップの高い領域に実注視点が多く存在すると考えられるため、2つの変量を元に正準相関分析を行うことで、双方の特徴の相関を高くする重みパラメータを得る。

2. 提案手法

図1は、提案手法の処理手順を示す。以下では、図1に従って、各処理の概要を順に説明する。

2.1 顕著性マップの生成

画像データの視覚的特徴を示す顕著性マップは、複数の特徴マップの線形結合によって求められる[3]。図2は、顕著性マップの算出過程に関する概念図を示す。まず、画像データ(original frame)に対して、6つの物理的特徴それぞれに関する画像解析を行った特徴マップ(feature map)を生成する。物理的特徴には、色(Color)、明度(Intensity)、方位(Orientation)、コントラスト(Contrast)、点滅(Flicker)、運動(Motion)がある。次に、各特徴マップの線形和が、顕著性

[†] (株) KDDI 研究所, KDDI R&D Laboratories, Inc.

[‡] NHK 放送技術研究所, NHK Science & Technology Research Laboratories

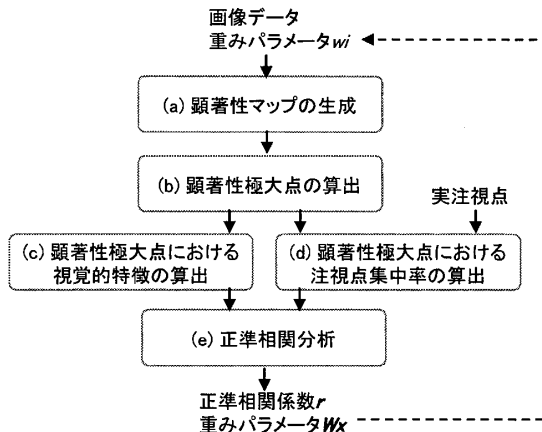


図1: 提案手法の処理手順

マップ(saliency map)である。図2の特徴マップおよび顕著性マップにおいて、白の濃い部分ほど顕著性が高い。

以下に、顕著性マップの導出式について示す。以下では、色、強度、方向、(輝度)コントラスト、点滅、運動のそれぞれの物理的特徴に関する識別子*i*を、*C*、*I*、*O*、*R*、*F*、*M*とする。また、各物理的特徴に関する特徴マップを FM_i とし、顕著性マップ SM に関わる各特徴マップの重みパラメータを w_i とする。つまり、顕著性マップ SM は、式(1)で表される。

$$\begin{aligned}
 SM &= \sum w_i \cdot FM_i \\
 &= w_C \cdot FM_C + w_I \cdot FM_I + w_O \cdot FM_O \\
 &\quad + w_R \cdot FM_R + w_F \cdot FM_F + w_M \cdot FM_M \quad (1)
 \end{aligned}$$

オリジナルのIttiらの顕著性マップ[3]では、 w_i を等重みとしている。また、特徴マップ行列データの各要素の値は $[0..MAX]$ の範囲 (MAX は特徴マップ行列データの要素の最大値)にある。

提案手法では、図1(a)において、動画コンテンツから切り出した画像フレームデータに対して、“Graph-Based Visual Saliency”ツール[10][11]を用いて、顕著性マップを算出する。ただし、従来手法のように w_i を常に固定するのではなく、後述の正準相関分析による重みパラメータ最適化によって得られた w_x の各要素を w_i に代入し、顕著性マップを算出し直す繰り返し処理を追加した。ここで、画像データ IM は、縦解像度 row 、横解像度 col のとき、以下の行列データで表すことができる。なお、行列 IM の要素はRGBの3値である。

$$IM = \begin{bmatrix} IM[1,1] & IM[1,2] & \dots & IM[1,col] \\ \vdots & \vdots & \ddots & \vdots \\ IM[row,1] & IM[row,2] & \dots & IM[row,col] \end{bmatrix}$$

これに対して、顕著性マップは、行数 row 、列数 col の行列データ SM として求められる。

$$SM = \begin{bmatrix} SM[1,1] & SM[1,2] & \dots & SM[1,col] \\ \vdots & \vdots & \ddots & \vdots \\ SM[row,1] & SM[row,2] & \dots & SM[row,col] \end{bmatrix}$$

以上より、顕著性マップ行列データ SM の要素 $sm = SM[p,q]$ ($1 \leq p \leq row, 1 \leq q \leq col$)についても、式(1)と同様に、式(2)で表される。

$$sm = SM[p,q] = \sum w_i \cdot FM_i[p,q], \quad i = C, I, O, R, F, M \quad (2)$$

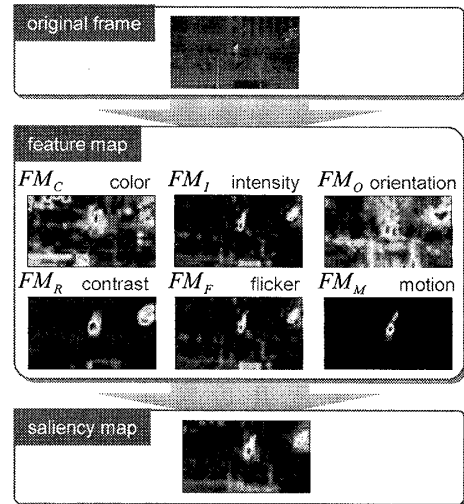


図2: 顕著性マップの算出手段

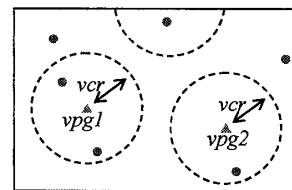


図3: 顕著性極大点の算出

2.2 顕著性極大点の算出

顕著性マップは、画像領域内での人間の注意の向けられやすさの分布を示したものである。つまり、顕著性マップの中で、顕著性の高い位置は、人間の注意が向けられやすい位置であることを示す。

提案手法では、図1(b)において、顕著性マップの行列データ SM の全要素の中で、 $SM[p,q]$ 値の大きい要素から順番に、「顕著性極大点」として抽出する。ただし、図3に示すように、各顕著性極大点同士(図3の三角印)は、一定距離 $2 \times vcr$ 以上離れているとする。これは、後述の顕著性極大点を中心とする中心窩領域(図3の点線円)を考慮するためである。

2.3 顕著性極大点における視覚的特徴

提案手法では、図1(c)において、顕著性極大点に対する「視覚的特徴ベクトル」を求め、後述する正準相関分析への入力パラメータ(変量群)の1つとする。

式(2)から、顕著性極大点 vpg に対する顕著性の値 $SM[p,q]$ は、各特徴マップにおける特徴量 $FM_i[p,q]$ の線形和で得られる。そこで、視覚的特徴ベクトルは、顕著性極大点における6つの物理的特徴量を要素に持つベクトルで表現する。つまり、顕著性極大点 vpg における視覚的特徴ベクトル $Vf(vpg)$ を、以下のように表す。

$$Vf(vpg) = [FM_C \quad FM_I \quad FM_O \quad FM_R \quad FM_F \quad FM_M]$$

さらに、顕著性極大点を中心とする中心窩領域内に、実注視点が含まれるように、多数の顕著性極大点を抽出する。つまり、 VG 個の顕著性極大点を抽出して得られた視覚的特徴ベクトルのデータセット VF を求める。

$$VF = [Vf(1)^T \quad \dots \quad Vf(VG)^T]$$

2.4 顕著性極大点における注視点集中度

提案手法では、図 1(d)において、顕著性極大点における「注視点集中度」を求め、後述する正準相関分析への入力パラメータ(変量群)の1つとする。

注視点集中度を求めるために、視線計測実験結果から、顕著性極大点を抽出した画像データに対応する実注視点を利用する。

人間には、視線方向の一点だけでなく、その周辺に高い精度で知覚できる領域(中心窩領域)が存在する。また、測定時の誤差も存在する。そこで、提案手法では、中心窩領域および測定誤差を考慮し、顕著性極大点と実注視点との一致の度合いを示す注視点集中度は、顕著性極大点を中心とする円領域に含まれる実注視点数の割合で評価する。つまり、顕著性極大点 vpg における注視点集中度 $Vg(vpg)$ を、以下のように表す。なお、顕著性極大点 vpg を中心とする半径 vcr の円領域内に存在する実注視点(図 3 の丸印)の数 $NS(vpg)$ 、実注視点の総数 NA とする。

$$Vg(vpg) = NS(vpg)/NA$$

さらに、2.3 節と同様に、 VG 個の顕著性極大点に対する注視点集中度のデータセット VG を求める。

$$VG = [Vg(1)^T \dots Vg(VG)^T]$$

なお、文献[12]では、眼球の網膜偏心率が大きくなるにつれ、視力が低下する視力分布の概念が示されている。また、文献[9]では、人間の注視点領域を、注視点を中心とする円領域の直径の両端と瞳孔とを結ぶ 2 本の線の角度が 2 度となる領域を考慮したガウスフィルタを畳み込むモデルが提案されている。提案手法は、これらの概念を導入することも可能である。

2.5 正準相関分析

2.5.1 正準相関分析の概念

正準相関分析は、2 組の変量群の間の相関関係を分析したい場合に利用され、2 つの変量群について、線形合成変数を求め、2 つの合成変数の相関が最も高くなるようにするものである[13][14][15]。このとき、正準相関分析の入力パラメータ(変量群)は、そのデータの特性を十分反映するように選ばれていなければならない[15]。

[13]では、Matlab 用関数 cca がある。これは、 $l = \min(m, n)$ として、 m 次元のデータを k 個集めた $m \times k$ 行列 X と、 n 次元のデータを k 個集めた $n \times k$ 行列 Y に対して、以下の式(4)によって $m \times l$ 行列 Wx 、 $n \times l$ 行列 Wy 、 $l \times 1$ 行列 r が得られる。

$$[Wx, Wy, r] = cca(X, Y) \quad (4)$$

ここで、 r は、2 組の変量群の相関関係(正準相関係数)であり、 Wx と Wy はそれぞれ、相関が最大になるときの、行列 X と Y に関わる重みパラメータである。

2.5.2 正準相関分析による重みパラメータの最適化

提案手法では、顕著性マップから抽出すべき情報と重みパラメータを同時に最適化するために、(1) 重みパラメータ w_i を固定して得られた顕著性マップから顕著性極大点を求め、(2) 顕著性極大点における重みパラメータを正準相関分析で最適化する、という操作を繰り返す。

つまり、図 1(e)において、視覚的特徴ベクトルのデータセット VF と注視点集中度のデータセット VG に対して正準相関分析を行い、正準相関係数 r および重みパラメータ

Wx を求める。さらに図 1(a)に戻り、求めた Wx の各要素を w_i に代入する。以降、(a)から(e)の処理を繰り返し行いながら、 w_i を最適化する。

3. シミュレーション実験

本章では、提案手法の効果を評価するために行ったシミュレーション実験について説明する。

3.1 視線計測実験

本シミュレーション実験では、先行研究[7][8]で行った視線計測実験で得た実注視点データおよび動画画像コンテンツを利用した。

視線計測実験では、瞳孔角膜反射法で視線計測を行った。視聴者は、プロジェクタからスクリーン(サイズ:100inch)に投射された動画画像(解像度:1920×1080)を、所定距離(約 3m)だけ離れた座席に着席して看視した。映像の画角は水平約 41 度、垂直約 24 度であった。

実注視点は、合計 80 名の被験者(男性 30 名と女性 50 名、21 歳から 48 歳、平均年齢 29 歳)から計測した視線データを元に計算した。つまり、動画画像コンテンツの再生開始時刻を開始点として、動画画像コンテンツの 150 枚の画像フレーム表示タイミングに合わせて計測した。ただし、瞬目などの影響で画像データ表示領域を外れた注視点は排除した。

動画画像コンテンツは、再生時間長が短く音声のない映像である。つまり、ドラマ、風景、CG、アニメなど、実際に放映されたテレビ番組の映像から、5 秒 150 フレームからなるクリップ動画画像である。短時間で音声のないクリップ動画を採用したのは、1 章で示したトップダウン要因を可能な限り排除できるように、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計測するためである。

3.2 解析データ

本シミュレーション実験では、前節の 15 種類の動画画像コンテンツの中から抽出した合計 50 枚の画像フレームデータからなる 3 種類のデータセットと、それらに対応する実注視点を解析対象とした。

以下にデータセットの求め方を示す。全動画画像コンテンツについて、150 枚の画像フレーム全般に渡る動き変化量の大小関係を解析した。つまり、それぞれの動画画像コンテンツに対して、コンピュータビジョン分野で利用される映像解析ツール MP-Factory[16]を利用した動き特徴に関する解析を事前に行った。動画画像コンテンツを構成する 150 枚の画像フレームに対して、まず、1 枚の画像フレームにおけるマクロブロックごとの垂直方向、水平方向の動きベクトルを求め、次に、全マクロブロック分のそれぞれの動きベクトルの標準偏差を求めた上で、全画像フレーム分の動きベクトルの和を求めた。

上記に基づき、本実験では、以下の 3 種類のデータセットを用意した。以下では、それぞれについて説明する。

[データセット 1 (DS-1)]

15 種類の動画画像コンテンツに対応する全画像フレームの中から、計測された実注視点数が最も多かった画像フレームから順番に 50 枚を抽出した。

[データセット 2 (DS-2)]

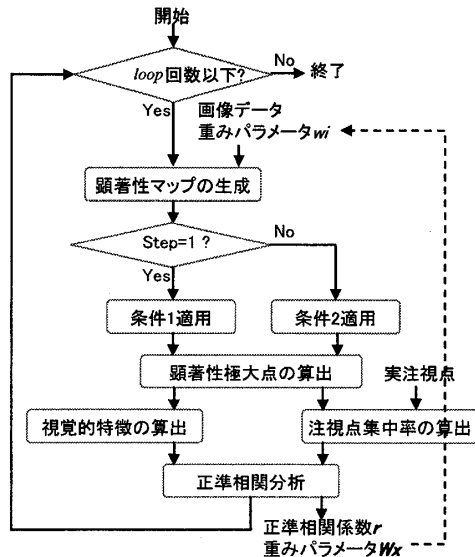


図4: シミュレーション実験処理手順

映像解析ツールによる解析結果から、動き変化量が最小だった動画コンテンツを1つ選択し、それを構成する画像フレーム150枚のうち、1枚目から50枚を抽出した。[データセット3 (DS-3)]

映像解析ツールによる解析結果から、動き変化量が最大だった動画コンテンツを1つ選択し、それを構成する画像フレーム150枚のうち、1枚目から50枚を抽出した。

3.3 解析方法

事前のシミュレーション実験において、顕著性マップから顕著性極大点を求めたところ、各顕著性極大点を中心とする円領域内に実注視点数が全く存在せず、2.4節の注視点集中度 $Vg(vpg)=0$ となる例が多く存在することが分かった。顕著性極大点の中には、視線(実注視点)を集める極大点と、集めない極大点とが存在するということである。顕著性極大点の位置は、顕著性マップ算出時の重みパラメータの設定によって変動する性質を持っているため、実験を行うにあたって、重みパラメータの初期値設定方法は重要である。実注視点を集める領域の推定に効果的な正準相関分析を行うためには、注視点特徴ベクトルの各要素が、0より大きい有効な値を示すことが望ましい。そこで、本実験では、顕著性極大点の算出および距離関係の分析を2段階で行うこととした。図4に、処理の流れを示す。

シミュレーション実験では、Step1とStep2の2段階から成る。それらはともに、2章で説明した手順に基づく。つまり、(1)重みパラメータ w_i を設定した上で、顕著性マップを生成し、顕著性極大点を求める、(2)(1)で得た顕著性極大点と実注視点に基づいて2つの特徴ベクトルを求めた上で、正準相関分析を行い、相関係数 r および重みパラメータ W_x を求め、(1)の w_i に代入する、という一連の処理を繰り返し行うことで、 w_i を最適化する。ただし、Step1とStep2は、顕著性極大点を算出するときの条件だけが異なる。Step1における条件1は、顕著性マップに対して、各顕著性極大点同士が距離 $2 \times vcr$ 分離れており、かつ、いずれかの実注視点との距離が vcr 以下であることを条件に、顕著性極大点を求めた。つまり、全顕著性極大点に対して、注視点集中度 $Vg(vpg)$ は非ゼロとした。一方、Step2におけ

る条件2は、顕著性マップに対して、各顕著性極大点同士が $2 \times vcr$ 分離れていることのみを条件に、顕著性極大点を求めた。

なお、視線計測実験では、解像度 1920×1080 の動画コンテンツを利用したが、本実験では、計算時間短縮のため、解像度 320×180 に縮小した画像データを利用して顕著性マップを算出した。顕著性マップを算出するために、[11]のツールを利用した。本ツールでは、[3]に基づく処理が行われるよう各種パラメータを設定した。

また、前記解像度の縮小に沿うように、実注視点データも座標変換を行った上で、顕著性極大点と実注視点間の分析を行った。

さらに、顕著性マップから顕著性極大点を抽出するとき、顕著性極大点同士は、解像度 1920×1080 の画像フレーム領域上で距離100ピクセル分以上離れていることとした。つまり、2.3節および2.5節の半径 $vcr=50$ (視角約1度)と設定した。

顕著性極大点数 Vg は、Step1のときは最大20個とした。また、Step2のときは、順番に求めた顕著性極大点に対して $\sum Vg(vpg)=1$ となるまで、抽出した。

本実験開始時の重みパラメータの初期値は、すべての i に対して $w_i=1$ とした。Step1では $loop=30$ 回繰り返し、Step2では $loop=10$ 回繰り返した。

3.4 実験結果

3.4.1 従来手法での顕著性と注視点集中度の相関

3種類のデータセットに対して、等重みで線形和を行う定義[3]どおりの顕著性マップから得られた顕著性極大点における、顕著性(重み付け和)と注視点集中度との相関分析を行った。つまり、重みパラメータの初期値を、すべての i に対して $w_i=1$ として、図1の(a)から(d)までの処理を行うことで、2つの値を求めた。これにより、すべての顕著性極大点に対して、顕著性 $SM[p,q]$ と、注視点集中度 $Vg(vpg)$ が得られるため、それぞれを軸とする散布図を作成した。

図5に、結果を示す。横軸は顕著性、縦軸は注視点集中度を示す。各散布図における相関係数は、DS-1の場合は0.0364、DS-2の場合は0.105、DS-3の場合は0.184であった。

3.4.2 正準相関係数の変化

図6に、3種類のデータセットに対して、提案手法のStep1およびStep2によって得られた相関係数の推移を示す。各グラフの横軸は繰り返し回数 $loop$ 、縦軸は相関係数 r を示す。

図から、Step1のDS-2については、 r に単調増加性と収束性があることが分かった。 r の最適化と重みパラメータの最適化を同時に行う提案手法の効果が現れている。しかし、それ以外のデータセットについては、そのような単調増加性および収束性が見受けられなかった。また、いずれのデータセットも、図5の結果よりも上回ったが、0.4から0.5の相関係数に留まった。さらに、Step=1のDS-1やDS-3は、繰り返しごとに r の変動が激しい。これは、正準相関分析を行うたびに、重みパラメータが更新され、顕著性マップを生成し直すため、顕著性極大点の位置も移動するからと考えられる。

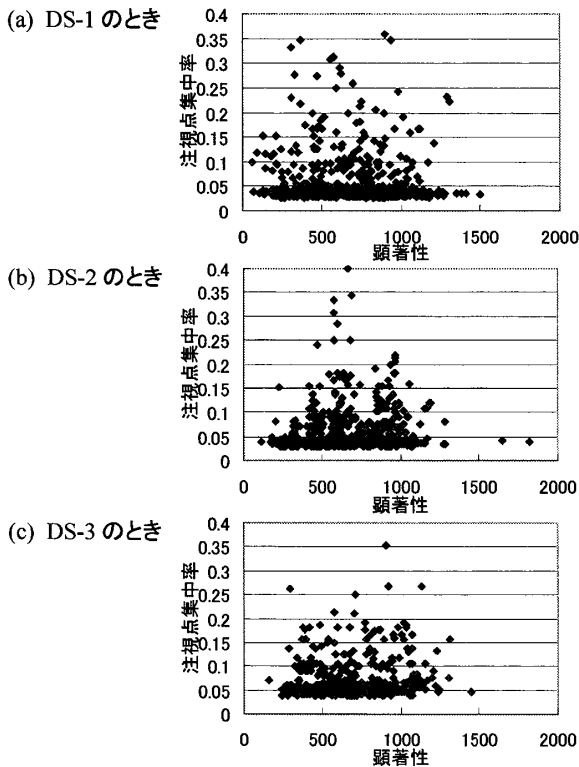


図 5: 従来手法における顯著性極大点の視覚的特徴量と注視点集中度との散布

3.4.3 顯著性極大点における顯著性と注視点集中度

図 5 と同様に、図 1(c)および(d)によって得られる 2 つの変量群に基づいて、各顯著性極大点に対する顯著性 (重み付け和) $SM[p, q]$ と注視点集中度 $Vg(vpg)$ を、繰り返しごとに求めた。図 7 に、顯著性極大点における顯著性と注視点集中度に関する散布図を示す。各グラフの横軸は顯著性、縦軸は注視点集中度を示す。なお各グラフは、Step2 で繰り返し回数 1 のときの結果を示したものである。各グラフの相関係数は、DS-1 の場合は 0.0597、DS-2 の場合は 0.297、DS-3 の場合は 0.161 であった。

3.4.4 動きパラメータの変化

図 8 は、正準相関分析により得られた、各重みパラメータの変動を示したものである。紙面の都合により、Step1 のときのみ掲載した。図 8 のそれぞれのグラフは、図 6(a) の 3 種類のデータセットに対応する。横軸は繰り返し回数、縦軸は各重みパラメータの値を示す。

図から、DS-2 において、Orientation の値が一番高いことが分かった。また、図 6 と同様に、DS-2 のときは、重みパラメータが収束していることが分かった。一方、DS-1 および DS-3 では、値の変動が大きい、Orientation が大きくなると相関係数が大きくなることが分かった。Motion の影響よりも Orientation の影響が大きく作用していることは、先行研究[7]の考察とも一致する。

4. 考察

本章では、前章のシミュレーション実験結果を踏まえて、提案手法について考察する。

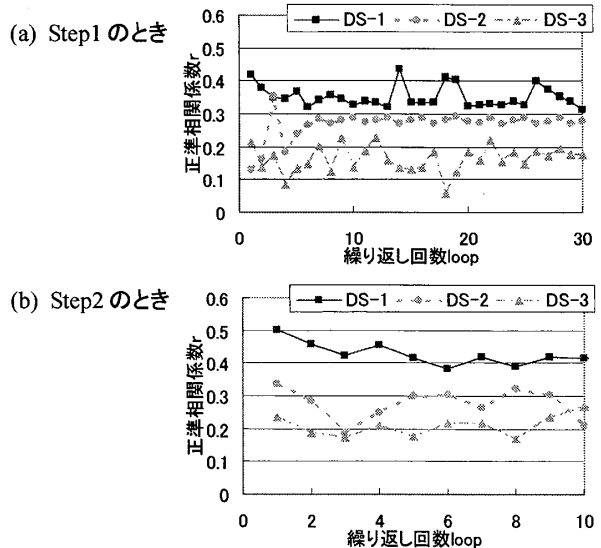


図 6: 相関係数の推移

4.1 動き特性と正準相関係数の関係について

今回の実験では、複数の動画画像コンテンツの画像データを混合させたデータセット DS-1 と、動き特性の異なる 2 種類のデータセット DS-2 (動き特性小) と DS-3 (動き特性大) を用いた。3.4.2 節で述べたように、正準相関分析を繰り返すことにより得られる相関係数は、動き特性の小さい動画画像の場合、収束し、3.4.1 節の定義に基づいて解析した結果よりも、高い相関があることが分かった。

実験では、サッカーボールの映像 (刺激が提示されてからサッカーボールが生じるまでの時間、200ms 程度) は考慮しなかった。そのため、実注視点と顯著性極大点の間には一定のずれが生じた可能性がある。また、3.2 節で記述したように、動き特性は、グローバルな動き成分を考慮したものであり、動き特性の大きい動画画像は、動き特性の小さい動画画像に比べて、顯著性極大点になり得る点が多くかつ画像領域全体に広がっている場合が多い。したがって、正準相関分析を行い、重みパラメータを更新し、顯著性マップを生成し直すごとに、顯著性極大点の位置も移動しやすくなる。その結果、動き特性の大きい動画画像の重みパラメータは、収束することなく、変動し続けたと考えられる。本実験では、実際にテレビ放映されたさまざまな映像を混合した解析を行ったが、今後は、1 つの動画画像の再生速度を大小変化させたときの注視点データと顯著性マップを元に解析することや、サッカーボールを考慮した場合、ローカルな動き特性を考慮した場合の効果についても検討する。

4.2 顯著性マップからの情報抽出精度について

図 7 において、縦軸 0.05 の付近に多数のサンプル点が存在している。これは、大多数の顯著性極大点の円領域の中には、少数の実注視点しか存在しなかったことを示す。視線の推定に役立たないこれら多数のサンプル点が、相関係数の推定に大きく影響していると考えられる。今後、注視点集中度を求めるにあたって、注視点推定に役立つ注視点マップからの情報抽出精度を高める方法を検討する必要がある。ちなみに、図 7 に対して、縦軸が 0.05 未満のサンプル点を除去した上で相関係数の変化を調べたが、法則性は見受けられなかった。さらなる検討が必要である。

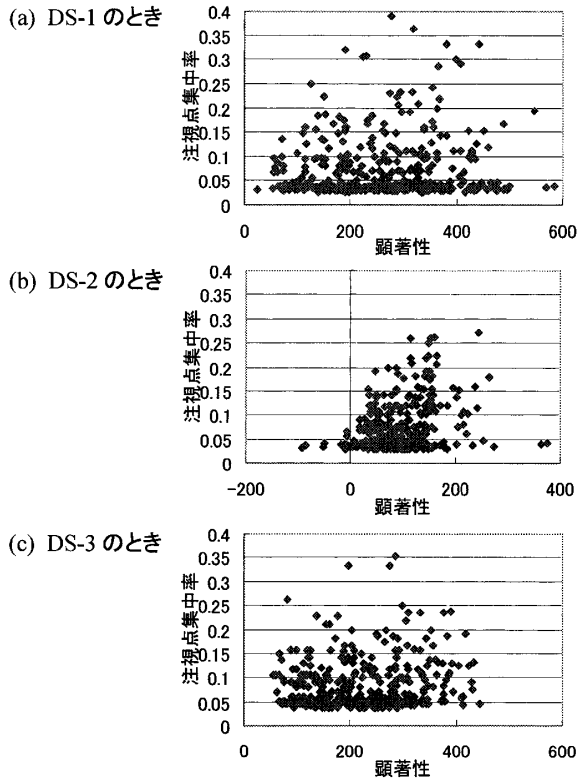


図 7: 提案手法における顕著性極大点の視覚的特徴量と注視点集中度との散布図(Step=2, 繰り返し回数=1)

また、実験では、トップダウン要因を排除できるような5秒の動画像を提示したが、顔や文字に視線が動く場合があったが、たとえ1秒の自然映像を用いても、トップダウンの影響が少なからず残ると考えられる。今後は、より単純な刺激映像を用いた実験も検討する。

5. おわりに

本稿では、動画像コンテンツの視覚的特徴を示す顕著性マップと、視線計測実験から得た注視点の分布との相関関係を正準相関分析によって分析し、顕著性マップ算出時に用いる重みパラメータを最適化する手法を提案した。提案手法は、膨大な情報量を有する顕著性マップと、多数の被験者から得た注視点データとを統計的に分析評価するところに特色がある。本稿を通じて、顕著性マップから、注視点の推定に役立つ局所的な情報を抽出し、実注視点との統計的分析によって、動き成分の少ない映像については、パラメータ最適化のプロセスが収束した。また、等重みで顕著性マップを求める従来手法よりも、高い相関が得られていることを示せた。

今後は、注視点推定に役立つ顕著性情報の抽出精度を高める手法を検討する。

参考文献

[1] 塩入 諭, 酒井 宏, 井上 哲理, 小峯 一見, 新井 啓之, “映像情報メディア年報 メディア情報 ヒューマンインフォメーション,” 映像情報メディア学会誌, Vol. 60, No. 8, pp. 1214-1220 (2006).
 [2] 澤島 康仁, 小峯 一見, 比留間 伸行, 伊藤 崇之, 渡辺 哲司, 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).

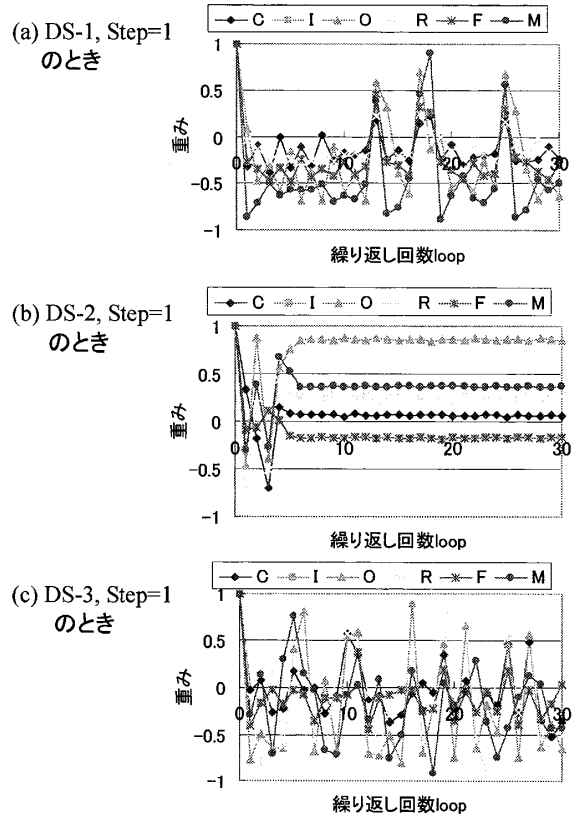


図 8: 重みパラメータの推移

[3] 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).
 [4] 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).
 [5] 三宅 洋一, 中川 慎司, “視線特性に基づく画質評価,” 電子情報通信学会論文誌 A, Vol. J89-A, No. 11, pp. 858-865 (2006).
 [6] 伊東 孝幸, 松宮 一道, 栗木 一郎, 塩入 諭, “視線位置の確率分布モデルを用いた誘目度の推定,” 第8回情報科学技術フォーラム (FIT2009), J-009 (Sept. 2009).
 [7] 上向 俊晃, 小峯 一見, 森田 寿哉, “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,” 第8回情報科学技術フォーラム (FIT2009), RJ-002 (Sept. 2009).
 [8] 小峯 一見, 澤島 康仁, 森田 寿哉, “多人数の視線を同時に測定する視線測定装置の開発,” HCG シンポジウム, HCG2009-B3-2 (Dec. 2009).
 [9] 中川 慎司, 津村 徳道, 中口 俊哉, 三宅 洋一, “注視領域推定を目的とした動画像特徴量の抽出,” 電子情報通信学会 2006 年総大会, AS-5-2 (2006).
 [10] J. Harel, C. Koch, and P. Perona. “Graph-Based Visual Saliency,” Proc. of Neural Information Processing Systems, pp. 545-552 (2006).
 [11] “Saliency Map Algorithm: MATLAB Source Code,” <http://www.klab.caltech.edu/~harel/share/gbvs.php>.
 [12] 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).
 [13] M. Borga, “Canonical Correlation Analysis,” <http://www.imt.liu.se/~magnus/ccal/>.
 [14] H. Hotelling, “Relations between Two Sets of Variates,” Biometrika, Vol. 28, pp. 321-377 (1936).
 [15] 柳井 晴夫, 他, “多変量解析実例ハンドブック”.
 [16] “MP-Factory Ver.4.1 (MPEG Software Development Kit),” <http://mmm.kddilabs.jp/ja/mpeg/mpfs/41/index.html>.