2008－AVM－60 (8)
2008／3／7

視線情報に基づく映像要約手法の評価

応和 大輔† 王 国明† 亀山 渉† 富永 英義†

† 早稲田大学 大学院国際情報通信研究科
〒 367-0035 埼玉県本庄市西富田大久保山 1011
E-mail: †owa@tom.comm.waseda.ac.jp, ††ong_kok_meng@fuji.waseda.jp,
†††{wataru,tominaga}@waseda.jp

放送のディジタル化や映像配信サービスの登場により，映像コンテンツはより身近なものとなった．膨大に増え続ける映像コンテンツの効率的なブラウジング技術のひとつに映像要約技術が存在するが，従来の要約手法の多くは映像の特徴や内容に依存しており，視聴するユーザの主観が考慮されているものは少ない．要約映像の品質はユーザの評価に依存するため，品質向上にはユーザの主観を無視することはできない．近年，瞳孔検出や視線検出の研究が進み，眼球運動計測装置が数多く登場している．眼球運動を計測することで人間がどこをどのように見ているのかを観測することができるため，コンテンツ評価の新たな指標として注目されている．本研究では眼球運動に着目し，ユーザの主観に基づく映像要約手法の検討と評価を行う．

Evaluation of Video Digest based on User's View Information

Daisuke OWA†, Ong Kok Meng†, Wataru KAMEYAMA†,
and Hideyoshi TOMINAGA†

† Graduate School of Global Information and Telecommunication Studies, Waseda University
1011 Okuboyama Nishi-Tomida Honjo-shi Saitama 367-0035 Japan
E-mail: †owa@tom.comm.waseda.ac.jp, ††ong_kok_meng@fuji.waseda.jp,
†††{wataru,tominaga}@waseda.jp

Digitalization of broadcasting and video delivery services has brought video content closer to people. There is video digest technique, which can manage the increasing number of video content efficiently. But the most of existing methods depends on content or feature in video, and tends not to take into account human subjectivity. Quality of digest video depends on user's feedback, so it is essential for quality improvement to consider the subjectivities. Recently, researches on detection pupil and gazing orbit have progressed, and many eye movement measuring devices become available. With those measurements of eye movement, we can observe where and how people watch, that may be used for content evaluation. In this paper, focusing on eye movement, a new method of video digest based on user's subjectivity is proposed and evaluated.

1. はじめに

近年，放送のディジタル化や映像配信サービスの登場，また，高性能PCの普及によって，ディジタルコンテンツに触れる機会が増えた．さらに，動画投稿・共有サービスによって，一般ユーザが制作した映像を広く一般に公開できるようになり，映像コンテンツは膨大かつ急激に増加している．しかし，映像コンテンツが増加し続ける一方で，我々が映像コンテンツを視聴するための時間は変わらず一定のままである．こうした背景から，我々の視聴スタイルは従来のながら視聴から，目的の映像コンテンツを選んで視聴するスタイルにシフトしつつある．映像コンテンツの効率的な管理と利用が求められる中，映像ブラウジング技術のニーズが高まりを見せている．

そのひとつである映像要約技術に関しては，ローレベルな符号的特徴およびハイレベルな内容情報を用いた知識のモデル化および内容の構造化によって，様々な映像のジャンルに対して要約手法が提案されてきた．しかし，作成された要約映像が真に映像コンテンツのダイジェストとなっているか，また，ユーザの欲する情報となっているかはユーザの評価による．そして，人間の趣味嗜好は千差万別であるゆえ，同じ要約映像でもその評価はユーザによって異なる．要約映像の品質向上にはユーザの主

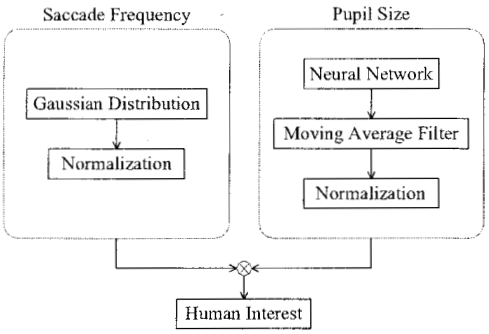

図 1　興味度算出フロー

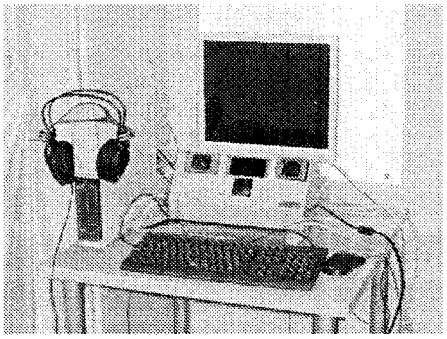

図 2　眼球運動計測装置[7]

観という要素を無視することは出来ず，また，この問題点は従来のアプローチでは完全に解決されていない．

そこで本稿では，ユーザの主観に基づく映像要約手法を検討する．

2. 生理指標としての眼球運動

心理学の分野では人間の情動反応を観測する方法として生理指標が用いられている．生理指標とは脳や神経の機能を人間の心理活動と関連付けることで，人間の内面の客観的な観測を可能とする手法である．生理指標は心理活動を物理量として定量化できるだけでなく，無意識領域をも実時間で観測することができるという特長があるが，反応までのタイムラグやノイズ，測定条件の制約が大きいなどの問題も抱えている．

生理指標を用いて観測された人間の情動反応を映像編集技術に適用する研究も既に行われている．豊沢は心拍変動情報のうち血圧性の低周波（LF）成分が人間の興奮と関連性があることに着目し，LF 成分の高い区間から心的負荷の少なくく，理解しやすい要約映像を作成している[1]．また，杉田らは血圧と心拍数の最大相互相関関数である ρ_{max} が精神的状況の変化を反映することを用いて，ユーザに提示する音声の音量をリアルタイムに調節するバイオフィードバックシステムを構築している[2]．

近年，数多くの眼球運動計測装置が登場しており，また，眼球運動の計測そのものがユーザの視聴環境に与える影響が小さいことから，本稿では，生理指標の中から眼球運動に着目し，人間の興味を検出し，これをユーザの主観として要約映像に適用する．生理指標としての眼球運動には以下の 3 つがあげられる．

- 瞳孔径　瞳孔の大きさと人間の興味度が比例する[3]．
- 視線の軌道　注視の傾向や対象に個人差が存在する[4]．また，情報の探索の際にサッケードが発生しやすくなる[5]．
- 瞬目　瞬目の頻度と人間の興味度が反比例する[6]．

以上のうち，本稿では瞳孔径および視線の軌道を用いて人間の興味を検出する．

3. 興味度算出アルゴリズム

本提案手法では，眼球運動計測装置から得られた映像視聴時のユーザの瞳孔径および視線の位置からシーンごとのユーザの映像に対する興味度を算出する．図 1 に興味度算出フローのブロック図を示す．

また，図 2 に本研究で用いた眼球運動計測装置の外観を示す．本装置は角膜反射法(注1)および暗瞳孔法(注2)を用いて，瞳孔および視点の位置を検出している．サンプリングレートは 60[Hz] で，瞳孔径やモニタ上における視点の絶対位置，計測の成否などを出力することができる．図 3. に計測装置から出力されるログファイルの内容の例を示す．

図 3. ログファイルの内容

```
⋮
00001386,01358829,0,L,04942,03542,0404,    F,     ,C
00001402,01358830,0,L,10240,07680,1000,   TF,XYP,C
00001417,01358831,0,L,05137,03470,0406,    F,     ,C
00001434,01358832,0,L,10240,07680,1000,P   F,XYP,C
00001450,01358833,0,L,05356,03323,0404,    ,      ,C
00001467,01358834,0,L,05444,03281,0404,    F,     ,C
⋮
```

3.1 瞳　孔　径

3.1.1 ニューラルネットワークによる補正

Hess の実験によって瞳孔の大きさが人間の興味の度合いに比例することが報告されている[3]．しかし，瞳孔はその径を調節することで，外界から眼球内に入射する光量を調節するという本来の生理的な機能も持っている．特に，強い光刺激によって瞳孔径が急激に変化する生理現象を対光反射と呼ぶ．図 4 に対光反射が発生した際の瞳孔径の変化を示す．瞳孔は眼球内に大量の光が入射すると（A），対光反射によって 0.2〜0.5 秒の潜時の後に縮瞳

(注1)：近赤外光を眼球に照射し，瞳孔および角膜表面における光源の反射像であるプルキニエ像を利用して視線を検出する手法
(注2)：瞳孔と虹彩の輝度差から瞳孔を検出する手法

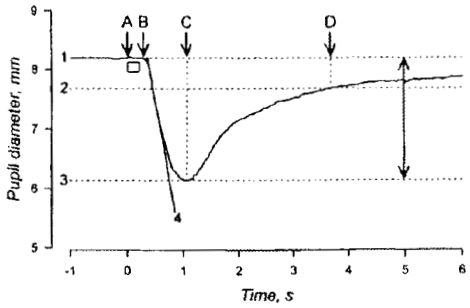

図 4　対光反射が発生した際の瞳孔径の変化[8]

を開始し（B），刺激提示から約 1 秒で最大縮瞳に達する（C）．瞳孔の大きさは心理活動のような内的要因と周囲の光量といった外的要因が合わさった結果であると言えるため，瞳孔径から人間の興味を検出するには対光反射の影響を除去する必要がある．映像の輝度は時間経過とともに変化し，常に光刺激が提示されている状態であると考えられるため，刺激提示から最大縮瞳に達する約 1 秒間の映像の輝度変化を考慮しなければならない．

この問題に対し，浅野らはニューラルネットワークを用いて瞳孔の大きさから対光反射の影響を取り除く手法を提案している[9]．表 1 に浅野らの提案するニューラルネットワークモデルの構造，図 5 に浅野らの提案するネットワークモデルのネットワーク図を示す．ネットワークに連続するフレームそれぞれの画面全体の平均輝度値を入力として与えると，相対的な瞳孔径が出力される．なお，入力の輝度値は YUV 空間における Y 成分の符号量を正規化した値，出力の瞳孔径は平均瞳孔径を 0.5 とした相対値である．

本提案手法では，上述の手法を用いて，計測装置によって得られた瞳孔径の値から映像の明るさによる対光反射の影響を除去し，補正する．表 2 に本稿で用いたニューラルネットワークのパラメータを示す．なお，学習回数は平均二乗誤差が収束すればよいものとする．ネットワークの教師データには[9]で瞳孔径の再現性が一番良好であった低輝度および高輝度の無地の映像が交互に出現する 0.5[Hz] の方形波の映像を視聴した際の計測結果を用いる．また，ネットワークの閾値および素子間の重みは学習ごとにランダムに与えているが，平均二乗誤差は常に 0.0002 付近で収束する．

このネットワークモデルでは，出力の相対瞳孔径は平均瞳孔径を基準としているため，教師データ用映像である無地のパターンの映像視聴時に心理活動がほぼ行われていないと仮定すると，平均瞳孔径とのずれが映像の明るさによる影響であり，補正すべき値となる．

3.1.2　移動平均フィルタ

次に，前述のニューラルネットワークを用いて補正した瞳孔径から高周波成分を取り除く．これは，瞳孔径の

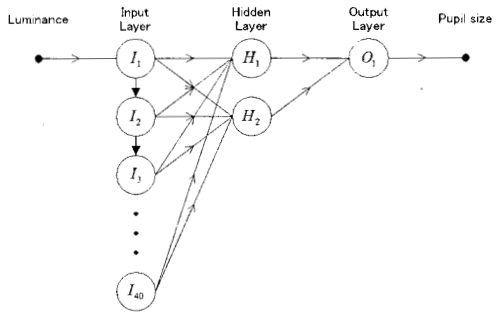

図 5　ネットワークモデルのネットワーク図

表 1　ニューラルネットワークモデルの構造

Learning Method	Back Propagation
Activating Function	Sigmoidal Function
Number of Layer	3
Number of Unit(INPUT)	40
Number of Unit(HIDDEN)	2
Number of Unit(OUTPUT)	1

表 2　ニューラルネットワークのパラメータ

Learning Counts	3000
Learning Ratio	0.05
Initial Entry(threshold)	-0.1∼+0.1(RN)
Initial Entry(weight)	-0.1∼+0.1(RN)

変動には固視微動などによる高周波の変動を含むためである．本提案手法では，式 1 で表される移動平均フィルタを用いて高周波成分を除去する．

$$P(i) = \frac{1}{M} \sum_{j=-L}^{L} P(i+j) \tag{1}$$

ただし，$P(i)$ はサンプル i における瞳孔径，M は平均化点数である．また，M と L の関係は以下の式 2 で表わされる．

$$M = 2L + 1 \tag{2}$$

この移動平均フィルタはローパスフィルタの一種であり，遮断周波数以上の高周波成分を除去することができる．平均化点数と遮断周波数の関係は以下の式 3 で求められる．

$$M = \frac{C}{f_c \Delta t} \tag{3}$$

ただし，C は利得が-3[db] となる場合の定数，f_c は遮断周波数，Δt は標本間隔である．なお，以下に本稿で用いた定数を示す．

$$\begin{aligned} C &= 0.442999 \\ \Delta t &= \frac{1}{60} \end{aligned} \tag{4}$$

ここで，ある被験者の複数の映像視聴時における瞳孔径の変動の周波数成分を図 6 に示す．スペクトルは低周

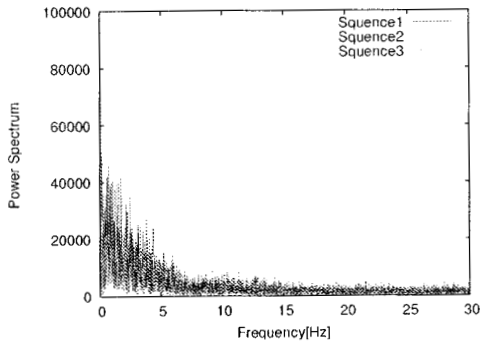

図 6　映像視聴時における瞳孔径の変動の周波数成分

波領域に集中しており，周波数の上昇につれ，スペクトルは減少傾向にある．また，5〜10[Hz] を境にスペクトルの減少傾向が緩やかになり，ほぼ一定になっているのがわかる．なお，被験者間およびシーケンス間で現象の差異は観測されなかった．

また，式 2 より平均化点数は必ず奇数である．そこで，式 3 を用いて求めた平均化点数が 3，5 および 7 の場合の遮断周波数を以下に示す．

$$f_c = \begin{cases} 8.860 & (M{=}3) \\ 5.316 & (M{=}5) \\ 3.797 & (M{=}7) \end{cases} \tag{5}$$

平均化点数が高いほど近似が強く，微小振動を除去できるが，図 6 からスペクトルは低周波領域に集中しているため，平均化点数が 7 の場合は情報の多くが欠落してしまう可能性がある．本提案手法では瞳孔径の変動から高周波成分を除去することを目的としているため，平均化点数が 7 以上の場合は適さない．また，平均化点数が 3 および 5 の場合はスペクトルの集中している低周波領域を保持することができるため，高周波成分を除去するためには平均化点数が高いものが望ましい．

よって，本提案手法では平均化点数を 5 に設定し，以下の式 6 を用いて 5.316[Hz] より高い周波数成分をノイズとして除去する．

$$P(i) = \frac{1}{5}\sum_{j=-2}^{2} P(i+j) \tag{6}$$

3.1.3　正　規　化

最後に，サッケード頻度から得られる興味度と尺度を統一するため，以下の式 7 を用いて正規化を行う．

$$P_s(i) = \frac{P(i) - \bar{P}}{\sigma_P} \tag{7}$$

ただし，$P_s(i)$ は標準化処理後の瞳孔径，$\bar{P}$ は瞳孔径の平均，σ_P は瞳孔径の標準偏差である．

3.2　サッケード頻度

3.2.1　正 規 分 布

サッケードは眼球運動が持つ代表的な 2 種類の機能である追跡と注視のうち，追跡の機能を持つ眼球運動であり，情報を探索する際に発生しやすい．前節でも述べたように，サッケードが頻繁に発生している際は，人間は視対象に集中し，情報の収集を行っている可能性が高い．また，サッケードは眼球運動の中でも最高速の運動であり，速度は最大で 600[°/sec] に達する[10]．

そこで，計測装置によって得られた視線の位置からサンプル間の移動量を算出し，適切な閾値を設定することで，サッケードのみを抽出することが可能であると考えられる．人間の生理的な特性として，サッケードは 1 秒間に 3 回までしか発生しないという報告があるため[11]，1 秒間ごとのサッケードの回数が 3 回以内に収まるような閾値を設定する．

しかし，サッケードの回数は 1 秒間ごとに算出しており，瞳孔径と異なり局所的かつ離散的な値をとるため，サッケードの発生によって興味度が急激に変化してしまう可能性がある．そこで，サッケードの回数に正規分布を適用することで幅を持たせ，興味度の変化を緩やかにする．本提案手法では，式 8 で表される標準正規分布を用いて，サッケードの発生前後 1 秒を含めることで，全体として 3 秒間の幅を持たせる．

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp^{-\frac{x^2}{2}} \tag{8}$$

3.2.2　正　規　化

最後に，瞳孔径から得られる興味度と尺度を統一するため，式 9 を用いて正規化を行う．

$$S_s(i) = \frac{S(i) - \bar{S}}{\sigma_S} \tag{9}$$

ただし，$S_s(i)$ は標準化処理後の i 秒時のサッケード回数，$\bar{S}$ はサッケード回数の平均，σ_S はサッケード回数の標準偏差である．

3.3　シーンごとの興味度

式 10 で表されるように，瞳孔径およびサッケード頻度から得られた興味度を加算し，シーンごとの映像に対する人間の興味度を算出する．

$$\textit{Interest Level} = \frac{1}{3.5 \times 60} \sum_{i=j\times 210+1}^{(j+1)\times 210} \left(P_s(i) + S_s(i)\right) \tag{10}$$

なお，興味度の高いシーン，すなわち重要度の高いシーンを検出する際に，興味度のゆらぎによってシーンが断片的になる可能性があるため，区間長処理を施す．人間の脳で処理できるシーンの最小区間長は 3.5 秒であると言われているため，本提案手法ではシーンの単位を 3.5 秒に設定する．計測装置のサンプリングレートが 60[Hz] で

あるため，本提案手法では 210 サンプルごとに平均した値を最終的な興味度とする．

本提案手法では，算出されたシーンごとの興味度を閾値によって興味のあるシーンとないシーンに分類し，興味のないシーンを早送りすることで要約映像を作成する．また，閾値と早送りの速度を変化させることによって，要約率を変化させることが可能となり，ユーザは要約映像の尺長を指定することが可能である．なお，閾値は尺長から自動的に決定される．要約率は以下の式 11 によって求められる．

$$\textbf{\textit{Digest Ratio}} = \frac{L_{above} + L_{below} \div PS_{fast}}{L_{all}} \times 100 \tag{11}$$

ただし，L_{above} は興味度が閾値以上のシーンの合計時間，L_{below} は興味度が閾値以下のシーンの合計時間，L_{all} は映像全体の合計時間，PS_{fast} は興味度が閾値以下のシーンの再生速度，すなわち早送りの速度である．また，各シーン長の関係は以下の式 12 で表わされる．

$$L_{all} = L_{above} + L_{below} \tag{12}$$

4. 評価実験

4.1 実験条件

本節では，前節で定義したアルゴリズムを評価すべく行った実験について述べる．本実験は被験者 5 名に対し，Video1 および Video2 の 2 件の映像を用いて，映像視聴時の被験者の眼球運動の計測結果から，本提案手法によってシーンごとの興味度を算出する．

また，本提案手法により算出された興味度と被験者の主観を比較するため，以下の側面から被験者にシーンごとの評価を行わせ，被験者の主観とした．

アンケート　映像視聴後に 5 件法を用いて評価を行う．前節で設定した映像の最小区間長である 3.5 秒ごとの評価は困難であるため，本実験では人手により分割した平均約 30 秒間のシーンごとに，以下の 5 段階で評価を行った．

- 興味がある（2）
- まあまあ興味がある（1）
- どちらでもない（0）
- あまり興味がない（-1）
- 興味がない（-2）

映像全体の評価が行える反面，評価の再現があいまいになる可能性がある．

ボタンデバイス　映像視聴中に図 7 に示すボタンデバイスを用いて評価を行う．被験者には興味のあるシーンでボタンを押すように指示をした．リアルタイムに評価が行えるが，部分的な評価しか行えない．

4.2 評価と考察

表 3 に Video1 におけるアンケートの一致率，表 4 に

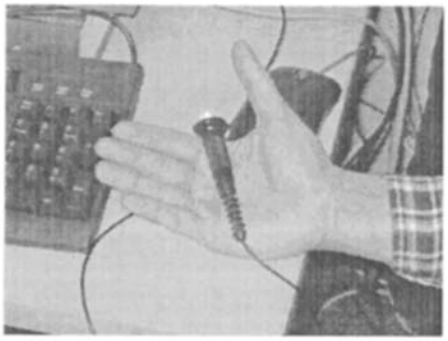

図 7　計測装置に付属するボタンデバイス

Video1 におけるボタンの一致率を示す．ここで一致率とは，主観評価によって興味ありと判定されたシーンにおける興味度が閾値以上である場合を正解とし，アンケートの場合は興味ありのシーンが占める割合を指す．本実験では，興味度の閾値を 0 に設定し，興味度が 0 以上のシーンを興味ありとする．なお，本実験において，アンケートによる評価が興味ありに偏ってしまったため，被験者ごとに評価の平均を算出し，平均以上の評価を得たシーンを重要度の高いシーンとした．表 3 は重要度の高かったシーンの一致率のみを示す．また，ボタンによって興味ありと判定されたシーンは，アンケートにおいても興味ありと判定されたシーンであることが多かった．

表 3 より，アンケートによる評価と一致率は平均で 5 割程度にとどまっており，比較的検出ができているシーンとそうでないシーンの二極化が起こってしまっている．本稿では，興味度の算出時におけるシーンの最小単位を 3.5 秒に設定したが，アンケートによる評価を行うにあたり，3.5 秒ごとの評価は困難であるため，人手により分割された約 30 秒程度のシーンごとに評価を行った．これにより，シーンの内部で興味度が推移してしまい，結果として一致率に影響を与えたと考えられる．また，シーン内部では興味度が上昇傾向にあるが，興味度の閾値まで達しない現象も確認されており，直前のシーンの興味度も影響していると考えられる．

表 4 より，ボタンによる評価と一致率は平均で 7 割であり，Subject3 を除けば比較的高い値となっている．また，不正解となったシーンに関しても，直前のシーンより興味度が上昇しているのが観測できた．なお，Subject3 は Video2 においても一致率が低く，生理指標は被験者の体調に左右されやすいため，生体反応が追従できていなかった可能性がある．

また，被験者の内省報告から，アンケートによる主観評価において，シーン全体の評価が局所的なシーンの印象に依存している箇所も確認され，アンケートによる評価における適切な区間長の設定も必要であると考えられる．

図 8 に Video1 における Subject5 の興味度と主観評価

表 3　Video1 におけるアンケートの一致率

	Scene1	Scene2	Scene3	Scene4	Scene5	Scene6	Scene7	Scene8	Scene9	Scene10	Scene11	Total
Subject1	70				41.18	51.04		38.24		77.78		54.49
Subject2	70	95.45	86.44	26.92	55.88					38.89		67.06
Subject3	70		70.34					36.76	15.63	25.93	40.70	46.96
Subject4	80	100									8.14	58.11
Subject5	80				41.18					90.74		47.42

表 4　Video1 におけるインタラクションの一致率

	Total Count	Accuracy Rate	Hit Rate
Subject1	4	4	100
Subject2	11	9	81.82
Subject3	8	3	37.5
Subject4	4	3	75
Subject5	4	3	75

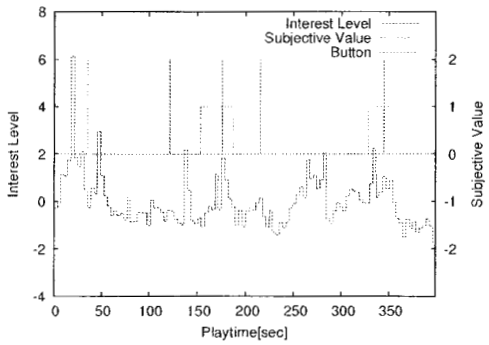

図 8　Video1 における Subject5 の興味度と主観評価

を示す．被験者がはっきりと興味ありと判定していないシーンにおいても興味度が高いシーンが存在する．これは生理指標が人間の無意識の領域をも観測できるためであると考えられる．

5. むすび

本稿では眼のふるまいから映像に対する人間の興味を検出し，要約映像に適用する手法について検討した．精度に関してはまだ課題が残されているが，算出された興味度と主観評価が一致したシーンも確認された．

今後の課題として，まず指標のロバスト性があげられる．本稿では，瞳孔径から対光反射の影響を除去する際に映像全体の平均輝度値を用いたが，映像のコントラストや色配置も考慮する必要がある．視線の動きに関しても，登場人物の会話のように音声情報が重要なシーンにおいては，サッケードは抑制され，話者を注視する傾向が観測されており，また，映像のパターンの持つ誘目性によって，無意識に視線が移動する現象も確認されているため[12]，映像のパターンによる注視傾向への影響を考慮する必要がある．本稿では用いなかった瞬目頻度や脈拍などの眼球運動以外の指標を取り入れることで，指標特有のノイズを抑え，興味の検出精度が上げることができると考えられる．

また，画像処理を用いた映像の内容や特徴を解析する従来の手法との協調処理を行って，精度向上を図る方法も考えられる．

謝辞　本研究は株式会社 VIS 総研との共同研究契約のもと同社の研究協力によって行われた．本研究を進めるにあたり技術的な協力を頂いた株式会社 VIS 総研に深く感謝いたします．

文　献

[1] 豊沢，河合，“心拍変動を利用した短縮映像作成方法，”ヒューマンインタフェース学会論文誌，Vol. 9，No. 2，pp. 243-249，May 2007.
[2] 杉田，田中，阿部，吉沢，山家，仁田，“情動反応を反映する生理指標の音響・映像を用いたフィードバック制御，”ヒューマン・インタフェース・シンポジウム論文集，Vol. 2002，pp. 125-128，Sep 2002.
[3] E.H. Hess，“Attitude and pupil size，”Scientific American，Vol. 212，pp. 46-54，Apr 1965.
[4] 松下，“テレビ画像を通して見る舞踊に関する眼球運動，”弘前大学教育学部紀要，Vol. 53，pp. 79-88，Mar 1985.
[5] 中村，井上，市村，松下，岡田，“「Ghost Tutor」眼球運動を利用した自主学習支援システム，”情報処理学会シンポジウム論文集，Vol. 2006，No. 4，pp. 93-94，Mar 2006.
[6] 田多，福田，山田，“まばたきの心理学 - 瞬目行動の研究を総括する，”北大路書房.
[7] “ナックイメージテクノロジー，”http://www.eyemark.jp/lineup/EMR-AT/EMR-AT.html
[8] P. BITSIOS，R. PRETTYMAN，E. SZABADI，“Changes in Autonomic Function with Age: A Study of Pupillary Kinetics in Healthy Young and Old People，”Age and Ageing，Vol. 25，No. 6，pp. 432-438，1996.
[9] 浅野，安池，中山，清水，“輝度変化に対する瞳孔面積変化モデル．”電子情報通信学会論文誌，Vol. 77，No. 5，pp. 794-801，May 1994.
[10] 福田，“生体情報システム論，”産業図書.
[11] Oliver Bergamin，Randy H. Kardon，“Latency of the Pupil Light Reflex: Sample Rate，Stimulus Intensity，and Variation in Normal Subjects，”Investigative Ophthalmology & Visual Science，Vol. 44，No. 4，pp. 1546-1554，April 2003.
[12] 橋本，牛木，中村，渡辺，小河原，“動画再生中における刺激提示の色の誘目性と配置に関する考察，”情報処理学会研究報告，Vol. 2006，No. 3(HI-117)，pp. 75-81，Jan 2006.