

受講者観察システムにおける動画像取得とサッケード検出の高速化

愛媛大学大学院理工学研究科数理物質科学専攻

森野聖治 †, 中川祐治 ‡

† s846009m@mails.cc.ehime-u.ac.jp , ‡ yuji@cite.ehime-u.ac.jp

1 はじめに

近年、一般家庭へのPCおよびブロードバンドの普及により、e-Learning、特にWebを用いた遠隔教育(WBT)が各方面で急速に用いられるようになってきた。しかし、受講者の状況を把握することが出来ず、各自の学習状況にあわせたコンテンツの配信順序・速度・内容などを提供できない欠点がある。そのため、学習を円滑かつ効率的に行わせるためには、臨場感を伴うインタラクティブな環境の提供と受講者の状況の把握、それに応じたコンテンツの制御が課題である。

我々は、上記課題の一つである受講者の状況を把握するために、受講者の眼球跳躍運動(以下Saccade)から学習状況を把握する受講者観察システム(Student Observation System)について考察してきた。このSaccadeは、読書時に必ず発生するものではなく、受講者が文章を読み理解している場合と、うつろな状態でただ文章を追っている場合とで、異なることがわかっている[1]。受講者観察システムとは、Saccadeを検知し、受講者の視線の動きを解析することで、学習状況を把握するシステムである。

本論文では、デジタルビデオカメラ(以下DVカメラ)を用いた映像信号の取得手法、黒目の中心座標を効率的に検出する手法、中心座標からSaccadeを検出する手法について述べる。

2 Saccadeについて

一般に人間が文章を読む際には、視野中心が文章を滑らかにトレースすることはなく、Saccadeと呼ばれる現象が発生する。Saccadeとは、視野中心が断続的かつ離散的な高速跳躍運動をする現象である[2]。Saccadeは1秒間におよそ3回、1回のSaccadeは約30msecで発生することが分かっている。

3 受講者観察システムについて

受講者観察システムは、受講者をDVカメラで撮影、顔画像から目領域を抽出、そこから黒目中心座標を検出、そして、人が文字を読むときに発生するSaccadeを検出し、受講者が文章を理解しているか否かを判断するものである[3]。また、Saccadeを検出するには、秒30コマのビデオレートでの撮影で、全てのコマについて観測することが必要となる。すなわち、1コマあたりの処理時間が33msec以内であることが要求される。

4 DirectShowを用いたDV信号の入力

DirectShowとは、Microsoft社のWindows用マルチメディア拡張API群であるDirectXに含まれるAPIの一つである[4]。DirectShowを用いることで、外部からのDV信号をPCに取り込むことができる。具体的には、DirectShowを用いてIEEE1394インタフェースから、映像信号をPCに取り込み、以下1~8の処理を施す。

1. システムデバイス列挙子の作成
2. デバイスの列挙
3. デバイスの決定
4. サンプルグラバフィルタの作成
5. サンプルグラバフィルタによる映像信号取得
6. 信号の2次元配列化(RGBに分解)
7. RGB成分を用いてGray Scale化
8. 5~7を繰り返しDV信号の取得を行う。

まず、システムデバイス列挙子を作成しデバイスを列挙する。そのデバイスをVideoInputDeviceCategoryで絞り込むことで、接続されているDVカメラを検出する。その後、サンプルグラバフィルタを作成し、映像信号を取得する。サンプルグラバフィルタは映像信号から画像を取得したあと、画像を変更せずにダウンストリームに渡す。サンプルグラバフィルタから取得された画像はG,B,Rの順にBitMap化される。G,B,Rの信号を用いてGray Scale変換し、メモリの2次元配列に保存する。

Gray Scaleへの変換方法は諸種あるが、ここでは、RGB成分に対してNTSC係数を用いた加重平均法と、G成分のみを利用したG-channel法を比較検討した。加重平均法では、輝度信号Yを

$$Y = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (1)$$

で表す。このYを用いてRGBをGray Scaleに変換する。G-channel法は、式(1)の係数の通りGにおける光の心理物理量が約6割ということを利用した方法で、輝度信号Yを $Y = G$ で表す。加重平均法の比率は、人間の目の感度に従っていて、より人間の感覚に近いGray Scaleに変換される。G-channel法は、計算式に処理が少ないため高速に処理できることが期待される。

5 黒目中心の検出

DV 信号を入力とした Gray Scale 画像から，テンプレートマッチング法 (以下 TM 法) を用いて目の部分のみを抽出し，その抽出された画像に対して以下 2~5 の処理を施し推定半径と黒目中心座標の決定を行う．

1. TM 法による目領域の抽出
2. 先鋭化 (*Sharpen filter*)
3. エッジ検出 (*Sobel filter*)
4. エッジ強調 (*Laplacian filter*)
5. 二値化
6. 黒目候補の決定
7. 推定半径の決定
8. 黒目中心座標の決定

二値化した情報をもとに視線の推移を検出し，Saccade の有無を判別する．

5.1 TM 法による目領域の抽出

TM 法とは，入力画像とテンプレート画像とを重ね合わせるにより比較照合し，両者が一致しているかどうかを判定する処理のことである．画像サイズの大きな入力画像からこの手法を用いて位置検出する場合には，画像サイズの冪乗に比例した計算コストがかかるので，SSDA (sequential similarity detection algorithm) 法を用いて高速化を検討した．本論文では，SSDA 法を変形させた以下の式を用いる．

$$m(u, v) = \sum_{k=1}^n \sum_{t=1}^m |g(u, v)| \quad (2)$$

ここで，高速化を図るため $m(u, v)$ が 閾値よりも大きくなった場合，処理を終了し次に移動する．全てのピクセル走査後の点 (x, y) における相違度を $D(i, j)$ ，最もテンプレートとの相違度が小さくなる点を $Min(g(x, y))$ とするとき，最適な目領域の切り出し始点 (i, j) は $D(i, j) = Min(g(x, y))$ を満たす (x, y) とする．ただし，上式を満たす (x, y) が複数存在する場合は，それぞれの平均値を与える．

5.2 入力画像の処理

目を含む領域画像に対して以下の式 (3)~(6) のフィルタ処理を施し，二値のエッジ画像を作成する．

5.2.1 先鋭化

画像前処理の第一段階として，画像に含まれるだけ成分を除去し，画像をくっきりと鮮明なものに変換する画像の先鋭化が考えられる．本論文では，既存の先鋭化フィルタを変形させた式 (3) のフィルタを用いる．先鋭化の重みを低くするために $w = 8$ とする．

$$S = \begin{bmatrix} 0 & -1 & 0 \\ -1 & w & -1 \\ 0 & -1 & 0 \end{bmatrix}, w = 8 \quad (3)$$

5.2.2 エッジ検出

上記処理において先鋭化された画像に対してソーベル・オペレータを施す．大きさが決められた近傍領域を用いて，より正確に水平・垂直方向の偏微分を計算するための方法として，近傍領域内の濃度の和を求めるときに上下左右の画素には対角線上のものより大きな重みを与える．水平方向，垂直方向のフィルタをそれぞれ式 (4) の H, V とする．

$$H = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, V = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (4)$$

フィルタ H, V を施した後の画素 (i, j) の値をそれぞれ $g_H(i, j), g_V(i, j)$ とすると式 (5) よりエッジ検出後の輝度値 $g(i, j)$ を得る．

$$g(i, j) = C_{Sobel} \cdot \sqrt{g_H(i, j)^2 + g_V(i, j)^2} \quad (5)$$

ただし， C_{Sobel} は定数とする．

5.2.3 エッジ強調

上記の処理で得られた画像に対して，更にエッジ情報を強調するために式 (6) に示すラプラシアンフィルタ L を施す．ラプラシアンフィルタとは注目画素 (i, j) の輝度値に対して 2 次微分を施すフィルタをいう．

$$L = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad (6)$$

5.2.4 二値化

最後にエッジ情報を含む画像を，閾値を用いた二値化処理によって二階調画像に変換する．本論文ではヒューリスティックに，二値化処理の閾値を決定している．入力画像の処理結果を図 1 に示す．



図 1: 目領域画像と入力画像の処理結果

5.3 推定半径と黒目中心座標の決定

黒目が円であると仮定し，“同一直線上に存在しない 3 点を通る円は唯一である” という公理より黒目を含むエッジ画像からエッジ情報をもとに円を発見し，黒目の中心座標を検出する．しかし，全ての組み合わせについて処理を行うと計算コストがかかりすぎてしまう．そこで，8 方向から順次画像を走査し，距離をもった 3 点の組み合わせを決定する手法を提案する．

5.3.1 円構成ピクセルの決定法

まず、画像を 8 方向に走査した集合を $D_p(p = 1, 2, \dots, 8)$ とし、図 2 の順に従い D_p を定義する。

例 [D_1 :左上隅から下に向かって水平方向右向きに走査]

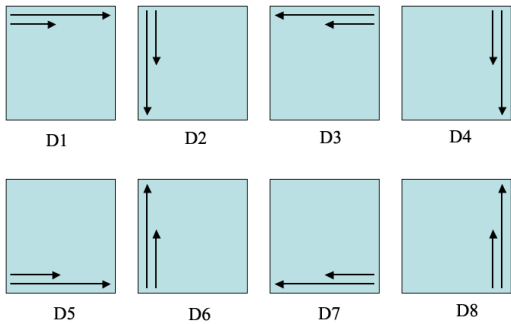


図 2: エッジの格納方向

エッジ画像からなる集合 D を、連続する 3 点が 1 つの円を形成すると考えられる組み合わせとなるように、3 つの集合 $E_i(i = 1, 2, 3)$ に分割する。分割した各集合 E_i の各要素に対して、以下のような一般項によって順番を付けると。

$$\begin{aligned} E_1(n) &= 3(n - 1) \\ E_2(n) &= 3(n - 1) + 1 \quad (n \in \mathbb{N}) \\ E_3(n) &= 3(n - 1) + 2 \end{aligned} \tag{7}$$

となる。ここで、エッジの総計を k とすると、各集合 E_i の末項 kk を以下のように定める。

$$kk = \begin{cases} k & \text{if } k = 0 \pmod{3} \\ k - 1 & \text{if } k = 2 \pmod{3} \\ k - 2 & \text{if } k = 1 \pmod{3} \end{cases} \tag{8}$$

これにより、番号が $(0, 1, 2), (3, 4, 5), \dots, (kk-2, kk-1, kk)$ のような順となる。

5.3.2 黒目候補, 推定半径, 黒目中心座標の決定

5.3.1 の 3 点に対し、条件 $(i_2 - i_1)(j_3 - j_1) \neq (i_3 - i_1)(j_2 - j_1)$ を満たすかどうかを判断し、これらを除く任意の第 4 点 (i_4, j_4) と中心座標 (a_l, b_l) に対し式 (9) を計算する。初回に限り D より上の条件を当てはめた 3 点にて中心座標を求めた後、式 (9) を計算する。

$$d_l = \sqrt{(i_4 - a_l)^2 + (j_4 - b_l)^2} \tag{9}$$

d_l が $|d_l - r_l| < \delta$ を満たすとき、アキュムレータ ACM_d (累算器) に累積する。また、このときの 3 点をアキュムレータ ACM_p に累積する。 ACM_p の累積値の平均を $E(ACM_p)$ とし、閾値 $E(ACM_p)$ より大きな値を持つ座標の集合を黒目候補点 \tilde{D} とする。 ACM_d は円の半径 r とその出現回数を表すヒストグラムであるとみなせるので、このヒストグラムの出現回数が最も多い値を \tilde{r} とし、 \tilde{r} を推定半径とする。 \tilde{D} に 5.3.1

を適用し、その 3 点の組み合わせに対して円の中心座標を計算し、アキュムレータ ACM_C に累積する。 ACM_C が最大となる座標を黒目中心座標とする。

6 Saccade の検出

これまでの研究 [5] では、全フレームに対して、目領域の抽出から黒目中心を求めるまでの処理を行ってきた。しかし、Saccade は 1 秒間におよそ 3 回発生することから多くのフレームは動きがなく計算結果が前フレームと同じになる、または、Saccade とは異なる動きをしている可能性が高い。つまり、ズレがないと判断した場合は黒目が動いていないとし、以降の各処理は全フレームの値を継承することにより処理の高速化を図る。また、受講者が読む文章は横書きのものとしているため、Saccade の検出には黒目中心の水平成分を用いる。もし、Saccade が発生すれば、その前後で黒目中心座標の水平成分が順方向に激しく変化する。

6.1 ズレの検出

二値化された 1 フレーム前の画像と現フレームの画像との差分をとり、差分値を d とする。 d を式 (10) の条件のもと解析し、ズレがあるかないかを判断する。

$$\begin{cases} \text{ズレ有り} & \text{if } d \geq \alpha \\ \text{ズレ無し} & \text{if } d < \alpha \end{cases} \tag{10}$$

ズレがあると判断された場合は、黒目の移動が主な原因であるため、Saccade が発生している可能性が極めて高いと考えられる。

6.2 処理制御

ここではビデオレートでの処理を確立するための一手法として、条件による処理制御 (表 1) を示す。

処理制御するための条件としてズレの有無と、前フレームで Saccade が発生したかどうか、という二つの条件を使用する。

表 1: 処理制御 (○:計算する ×:計算しない -:未定義)

Saccade	ズレ	推定半径	黒目候補	黒目中心	Saccade
-	-	○	○	○	×
有	有	○	○	○	○
	無	継承	×	継承	○
無	有	継承	○	○	○
	無	継承	×	継承	○

7 実験

ここでは 1 フレームあたり 33msec 以内での処理ができるかどうかを調べるため、DV 信号の入力、TM 法、入力画像の処理、黒目中心と半径の検出、Saccade の検出における処理速度の測定結果を記す。また、Saccade の検出結果を記す。

まず，DV 信号入力の実験として DV 信号の入力手法の処理時間を計測した．今回は入力信号として，G-channel 法，NTSC 加重平均法を用いた Gray Scale，RGB 値による color の 3 種類に注目し，それぞれの平均処理時間を測定した．3 種類それぞれにおいて，5000 回計測し平均値を求めた結果，表 2 のようになった．DV 信号の取得で使用した環境は，PC は OS が WindowsXP，CPU が CeleronM360 1.4GHz，カメラは SONY DCR-TRV310 である．

表 2: DV 信号取得の平均時間 [msec]

G-channel 法	加重平均法	Color(RGB)
8.9794	32.321	1.4648

次に，TM 法による目領域の抽出における実験として 5.1 で述べた手法の処理時間を計測した．DV カメラで取得した 10 フレームに対して TM 法による目領域の抽出に要する処理時間の計測結果は表 3 となった．フレーム No.4~6 は瞬き中により目領域の抽出はなかった．目領域の抽出，黒目の中心と半径の検出では，OS が MacOSX ver10.3.9，CPU1.25GHz PowerPC G4 の PC を使用した．

表 3: TM 法における類似度と処理時間

フレーム No	類似度 [%]	処理時間 [sec]
1	96.5	11.7027
2	90.6	0.6571
3	82.7	0.6189
4	0	0.5537
5	0	0.5451
6	0	0.5764
7	83.4	0.5937
8	91.6	0.6185
9	94.3	0.6023
10	94.5	0.6043

目領域画像に対する先鋭化から二値化までの処理時間の平均は，2.5msec となった．

また，目の中心と半径を実際に 5.3 の手法を用いた実験を行った．処理時間の平均を求めた結果，表 4 のようになった．所要時間は従来の約 1/180 となった．

表 4: 円の中心と半径の検出における処理時間 [msec]

高速化アルゴリズム	従来のアルゴリズム
0.3845	66.2735

最後に，ズレの検出の処理時間は 0.38msec，Saccade の検出の処理時間は 0.03msec となった．

Saccade の検出の実験として図 3 に対象文章を集中して読んでいるときの眼球の動きと疲労時における眼

球の動きを示す．集中時には 4 回の Saccade が検出された．疲労時には目が逆方向に動く場合もあり，うつろな状態であることがわかる．

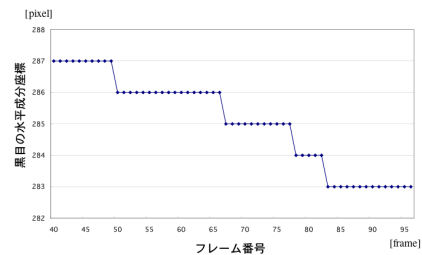


図 3: 集中時における眼球の動き

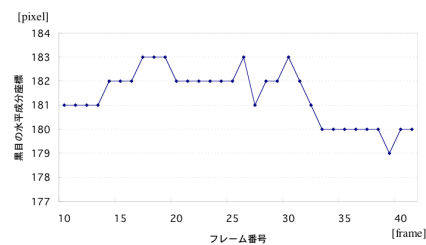


図 4: 疲労時の眼球の動き

8 まとめ

本論文では，受講者観察システムにおける DV 信号の入力，目領域の抽出，黒目中心と半径の検出の実験を行った．DV 信号に対し G-channel 法を用いた Gray Scale への処理時間は表 2 に示される様に 1 フレームあたり約 8.9msec と実用に耐えうる結果となった．また，黒目の中心，半径検出の処理時間も従来の 1/180 という結果となった．TM 法による目領域の抽出以外での合計時間は，約 12.5msec と範囲内の結果となった．Saccade の検出に関しては，集中時とうつろな状態での実験結果の通り，Saccade の有用性を再確認することができた．今後，目領域の抽出におけるアルゴリズムの高速化を考察するとともに，1 フレーム毎に全ての処理を行うことをめざす．

なお，本研究の一部は，平成 17 年度科学研究費補助金（基盤研究 (C)）No. 17500646 による．

参考文献

- [1] 加藤亮彦, 中川祐治, "受講者観察システムの研究-サッケード検出の高速化-", 電子情報通信学会技術研究報告, Vol.105, No.336, pp.75-78, 2005
- [2] 池田光男, "眼は何を見ているか", 平凡社, 1988
- [3] 中川祐治, "読書時の眼球運動を利用した受講者観察システム", 2003 年度人工知能学会全国大会, 2004
- [4] e-words : <http://e-words.jp/w/DirectShow.html>
- [5] 杉本圭, 中川祐治, "眼球運動の検出による受講者観察システム", 2003PC カンファレンス論文集, pp.219-222, 2003