

眼球運動に基づく動的フレームレートでの視覚情報の獲得

堀 泰浩*¹ 吉高 淳夫*¹

概要

本研究では、注目時に現れる眼球運動の特徴である、固視状態と跳躍運動の繰り返しを検出し、注目対象が静止していれば静止画像、視覚的に変化していればその変化に応じてフレームレートを変更した動画像として注目した視覚情報を獲得する。提案手法では、眼球撮影用のカメラよりユーザの眼球運動を検出する。そして、視界を撮影するカメラから得られる視界画像において有効視野領域の変化を検出し、注目対象の視覚的な変化の頻度から、静止画像、あるいは注目対象の変化を捉えるに十分なフレームレートの動画像として、視覚情報を獲得する。

Vision Acquisition with the dynamic frame rate based on Eye Movement

Yasuhiro Hori*¹ Atsuo Yoshitaka*¹

Abstract

Human's eyes are known to take a typical pattern under the situation of gazing visual information. In this paper, we propose a method of acquiring vision with the dynamic frame rate based on eye movement. When a user gazes at an object and the object is stationary, the proposed system acquires the vision as a still picture. And if the object is varying visually, it acquires the vision as a video which the object's visual variation is sufficiently recorded. This system detects user's eye movement with a camera placed below an eye and detects variation of the effective visual field with a camera which records the user's vision. Based on the frequency of the visual variation of the gazed object, the system determines the frame rate which the object's visual variation is sufficiently recorded.

1 はじめに

人間は、視覚から様々な情報を得ることによって見たことや体験したことを記憶している。しかし、人間の記憶に蓄えられた情報は、思い出そうとしても不正確であったり、時間の経過

とともに不鮮明なものとなっていく。そうしたことから、正確に記憶を呼び起こそうとしてもなかなか思い出せないといったことがある。そのため、人間は注目し、興味を示した情報をメモとして記録したり、写真を撮ったりすることによって、記憶の補助を行っている。しかしながら、このような方法では、明示的な動作が必

*1: 広島大学大学院工学研究科

*1: Graduate School of Engineering, Hiroshima University

要となり、その場で行っているユーザの作業や思考を一時中断させることになる。そのため、人間の記憶となる視覚情報を自動的に獲得し、提示することが可能となれば、記憶の補助に有効な手段となる。

記憶の補助を行った研究として[1]がある。人間の眼球運動は、視覚情報注目時に約 300 ミリ秒間の固視状態と約 30 ミリ秒間に起こる跳躍運動を繰り返すことが知られている[2]。図形・風景に注目しているとき、視線は図形のエッジ、境界、運動部分などの特徴点間で固視状態と跳躍運動を繰り返し、文章に注目しているとき、視線は文章に沿って固視状態と跳躍運動を繰り返す。滑らかに移動する動物体に注目しているときのみ、視線はスムーズに移動する。[1]では、このような注目時の眼球運動に着目して、ユーザが意識して注目した視覚情報を自動的に獲得し、情報注目時のユーザの状況と関連付けてデータベース化している。そして、スケジュール帳と連携した検索・管理インタフェースを構築することで容易な検索を可能とし、獲得した視界画像をユーザに提示することで記憶の補助を行っている。しかしこの手法では、視覚情報を主に静止画像として獲得し、注目対象が滑らかに動く動物体に限り動画像で獲得を行っているため、TV 画面や、文章が流れるように表示される電光掲示板などの、視覚的な状態が頻繁に変化する対象に注目している場合に、注目対象の情報を十分に獲得できない。本研究では、注目対象の視覚的な変化を動画像として記録することで、より記憶の想起に有効となると考え、視覚情報を注目対象に応じて静止画像、または動画像として獲得する。

仮に、注目した視覚情報を全て動画像として獲得した場合、必要となるストレージの量は膨大になり、また、視覚的な変化が見られない注目対象も動画像として獲得してしまうことで冗長性も大きくなる。このような問題を解決するために、ユーザが注目している対象が、視覚

的に変化しているかどうかを判断し、その変化の度合いに応じて視覚情報の保存形式を変更して獲得する。本研究では、視覚情報注目時において視線の跳躍が生じた際に、注視した部分での注目対象の視覚的な変化の検出を行うことで、視覚情報を静止画像、または注目対象の視覚的な変化を捉えるに十分なフレームレートによる動画像での獲得を行う。

2 関連研究

視覚情報を動画像として獲得する研究として、[3]-[5]が挙げられる。[3]では、頭部に装着したカメラより、ユーザが興味を示した視覚対象を動画像として獲得し、再び同じ対象を見たときに、過去に獲得した動画像をユーザに提示することで記憶の補助を行っているが、動画像の獲得には明示的な操作が必要である。[4]では、実世界のオブジェクトに RFID タグを取り付け、オブジェクトに触れることでその時の視覚情報を動画像として記録している。[3]、[4]の手法は、注目した対象が静止対象、動物体に係わらず、固定のフレームレートで獲得している。また、[5]では、頭の動きや、注目対象の動きに基づいて、獲得した日常行動の動画像を、注目した個所を抜き出すことによって要約する手法を提案している。この手法は、注目対象が静止しているか、動いているかの判断は行っているが、獲得方法は注目対象の視覚的な変化の有無に関係なく動画像として獲得している。本研究では、注目時の眼球運動を解析し、ユーザが注目した情報のみを自動的に獲得する。そして、注目対象が静止対象であれば静止画像、視覚的な変化のある対象であれば、その変化を十分に確認できるフレームレートの動画像として獲得することで、注目対象に応じた視覚情報の獲得を行う。

3 システム構成

システム構成を図1に示す。ユーザの眼球を撮影するCCDカメラはユーザの眼球の下部に固定する。また、ユーザの視界を撮影するCCDカメラはユーザの視界が得られるようにユーザの眉間に固定する。ただし、ユーザが正面を見ているときの眼球とその時の視界の中心がそれぞれのカメラ画像において中心に映るようにする。入力画像に対して画像座標系は図2のようになる。カメラを装着したときの様子を図3に示す。

眼球を撮影するカメラの画像から瞳孔の動きを検出し、ユーザが注目状態であるかを判断する。注目が開始した時点で、視界を撮影するカメラからの画像の有効視野領域において、注目対象の視覚的な変化を検出すると同時に、視界画像のバッファリングを行う。視覚情報への注目が終了した時点で、注目対象の変化の頻度に基づいてフレームレートを決定し、バッファリングした視界画像を、静止画像、あるいは動画像として視覚情報データベースに格納する。

現在CPUがPentium、500MHzのノート型PCを使用し、処理する映像データのフォーマットは、フレームサイズは2台のカメラとも160×120[pixel]、処理速度は10[fps]である。なお、眼球を撮影する映像は256階調グレースケール、視界を撮影する映像は24bit colorでキャプチャしている。

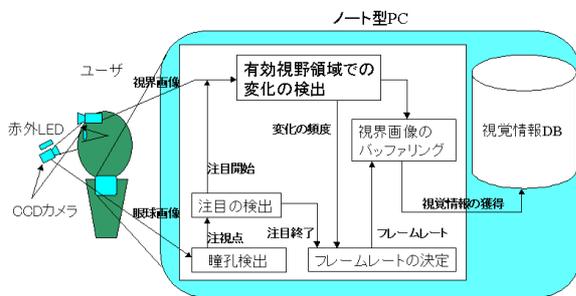
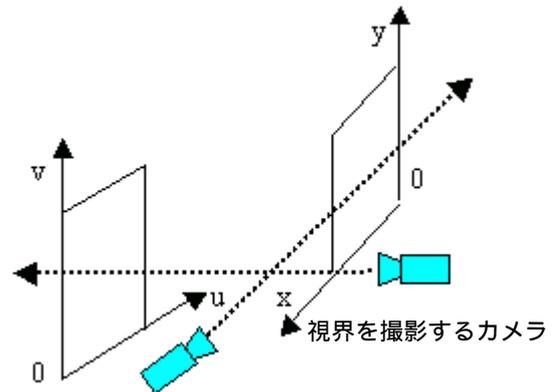


図1 システム構成



眼球を撮影するカメラ

図2 画像座標系

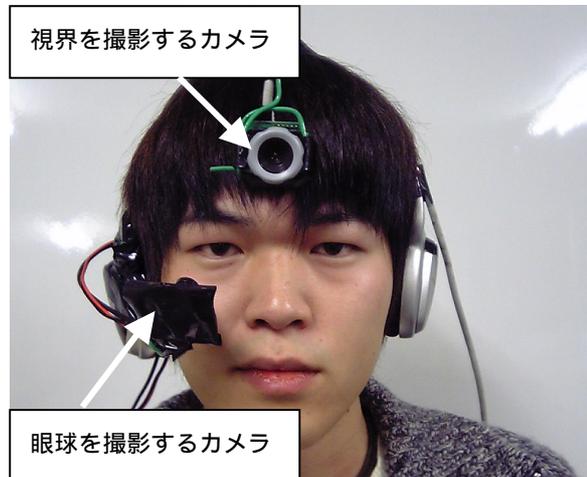


図3 カメラ装着時

4 注目対象の変化に応じた視覚情報の獲得

ユーザが周囲の視覚情報に注目していることを判断するために、眼球を撮影するカメラから視覚情報注目時の視線の動きの検出を行う。そして、注目対象が視覚的に変化しているかどうかを判断する。

まず、ユーザの瞳孔を検出し、ユーザの注視点座標を決定する。そして、視界画像から注視点を中心とした有効視野領域において、注目対象の視覚的な変化を検出し、変化に応じた視覚情報の獲得を行う。

4.1 視線の検出及び注視点座標の決定

・視線の検出

ユーザの視線の検出には、虹彩及び瞳孔の位置を検出することによって行う。正確な視線の位置を求めるため、眼球に赤外線を照射することで瞳孔と虹彩のコントラストを強調し、眼球画像の二値化処理によって、瞳孔部分のみを抽出する(図4、5)。そして、瞳孔領域の重心座標 $(x(t), y(t))$ を瞳孔の位置とする。ここで、 $x(t), y(t)$ は時刻 t (単位はフレーム) における瞳孔の位置の x, y 座標である。ただし、眼球画像の水平方向を x 軸、垂直方向を y 軸とし、原点を画像の左下とする。



図4 眼球画像

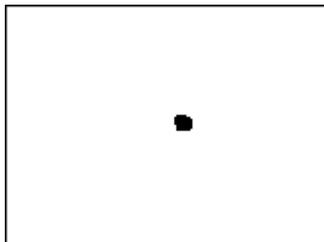


図5 瞳孔領域

・注視点座標の決定

注視点とは、ユーザが視界画像中のどこに視線を合わせていたかを示す座標であり、瞳孔の位置から求める。視界画像における水平方向を u 軸、垂直方向を v 軸、原点を画像の左下としたときの注視点の座標 $(G_u(t), G_v(t))$ を正面を見ているときの瞳孔の中心座標 (D_x, D_y) を用いて次の式によって求める。

$$\begin{cases} G_u(t) = Cv_x - (x(t) - D_x) \times \alpha \\ G_v(t) = Cv_y + (y(t) - D_y) \times \beta \end{cases} \quad (1)$$

ただし、 Cv_x, Cv_y は視界画像の中心座標とする。また、 α, β は瞳孔の中心の移動距離に対する注視点の比を表す値であり、今回は $\alpha = 7.5, \beta = 11.5$ とした。

ある点を注視したときの瞳孔領域とそのときの注視点の一例を図6、図7に示す。図中の十字は画像の中心を示している。

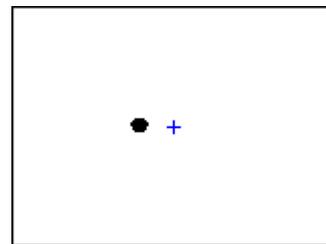


図6 瞳孔領域



図7 注視点

4.2 注目対象の視覚的な変化の検出

ユーザが視覚情報に注目する際、ユーザの体及び首は静止した状態であるとする。

まず、注目が開始したことを眼球運動の解析によって判断する。提案手法では、視線の跳躍が、1回の跳躍につき3秒以内の間隔で、3回以上検出される場合に視覚情報に注目しているものと判断している。ただし、瞳孔の位置が眼球を撮影するカメラにおいて3ピクセル以上移動した場合を視線の跳躍が生じたものとし、最初に視線の跳躍が生じた時を注目の開始とする。また、視線の跳躍が3秒間検出されなかつ

た場合、その時点で注目が終了したとみなす。

そして、注目の開始に伴い、視界を撮影するカメラからの画像をバッファリングすると同時に、注目開始時のフレームと、その後のフレームとの比較を行い、注目対象の視覚的な変化の検出を行う。注目が終了するまで、視界画像のバッファリング、及び有効視野領域における注目対象の変化の検出を行う。

・ 有効視野領域における変化の検出

注目が開始した時点の視界画像を基準フレームとし、その後の注目区間においてバッファリングされる各フレームと比較を行う。

基準フレームとその後獲得される各フレームとの比較は次のように行う。それぞれの視界画像における有効視野領域(図8)と、その領域と同座標に位置する基準フレームの領域に対して、各画素のRGB表色系における色差を求める(図9)。

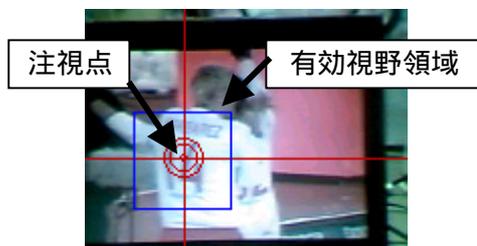


図8 有効視野領域

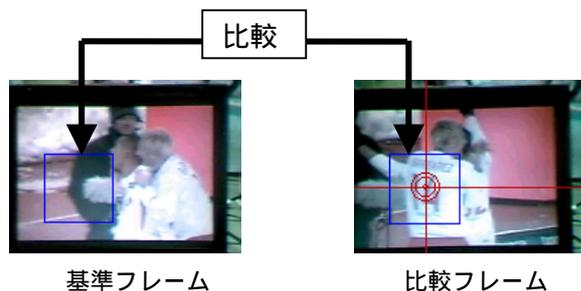


図9 領域間の比較

領域間の比較の結果、色差が閾値 Th_{cd} 以上の画素が比較領域の総画素数の閾値 $Th_{pn}\%$ 以上の割合を占める場合、ユーザの注目している対象

は視覚的に変化したと判断する。今回用いた各閾値は、実験により、提案手法において変化の検出に最適であると判断した、 $Th_{cd}=50$ 、 $Th_{pn}=30$ を用いた。注目対象が視覚的に変化すると判断された場合、その時の視界画像を新たな基準フレームに置き換え、次のフレームとの比較を行う。

4.3 フレームレートの決定

視覚情報への注目が終了した後、バッファリングした視界画像を注目対象の視覚的な変化の頻度に応じたフレームレートにより、動画像、あるいは静止画像として視覚情報データベースに保存する。フレームレート R は(2)式により決定する。ただし、注目を開始してから終了するまでに処理したフレーム数を F_n 、注目区間において、4.2.2で述べた注目対象の視覚的な変化の検出を行い、視覚的に変化すると判断された回数を N とする。また、 P_s は、本システムの処理速度[fps]であり、現在は $P_s=10$ である。

$$R = \frac{P_s \times N}{F_n} \quad (2)$$

獲得される動画像の最大フレームレートは10fpsであり、注目対象の視覚的な変化が激しい場合はフレームレートを高く、穏やかな場合はフレームレートを低くした動画像として視覚情報を獲得する。ただし、 $R < 1$ となる場合は、注目対象に視覚的な変化がほとんど生じなかったと判断し、静止画像として獲得する。

5 一定方向へ移動する対象に注目した場合の視覚情報の獲得

スクロールする文字列である電光掲示板等は、同一色の背景に単色の文字情報が一定方向に移動しながら表示されている。そのため、視覚的な変化の検出に色差を用いた前述の方法では変化が検出されにくく、静止画像として獲得してしまい情報を十分に得られない。このよ

うな場合は、注目時の視線の移動方向に着目して、ユーザが一定方向への移動を行う対象に注目していることを判断することで、視覚情報を動画像として獲得する。

静止対象である絵画画像に注目した場合の視線の動き(図 10(a))と比較して、電光掲示板などの限られたスペースでスクロールする文字列に注目する場合、静止対象に注目している場合と異なり、特徴的な眼球運動が見られる。図 10 の(b)に示すように、視線は水平方向に連続して移動しており、垂直方向への視線の移動はほとんど見られない。ただし、横書きの文章に注目している場合でも、文字に沿って視線の跳躍が生じるため、視線は水平方向に連続して移動する。しかし、文章に注目している場合は、下方に読み進めていくために、視線の移動には垂直方向への移動が含まれることから、視線の移動方向が水平方向に連続して現れ、なおかつ垂直方向への移動がない場合に、電光掲示板のような一定方向に移動する対象へ注目していると判断することができる。水平方向への移動とは、視線の跳躍が生じた際の、跳躍前のフレームと跳躍後のフレームにおける注視点座標を結ぶ直線の傾き a が $|Th_h|$ 以下である場合とする。また、垂直方向に移動する対象の場合は、傾き a が $|Th_v|$ 以上となる場合が連続し、水平方向への視線の移動がない場合に、垂直方向に移動する対象に注目していると判断する。今回用いた閾値は実験により最適と判断した $Th_h=0.50$ 、 $Th_v=2.00$ とし、水平方向、あるいは垂直方向の視線の移動が、3 秒間連続している場合に一定方向に移動する対象に注目していると判断する。

このような一定方向へ移動する対象を動画像として獲得する場合、視線の移動方向からは、注目対象の移動の速度が判断しにくく、フレームレートの決定が困難であることから、人間の

目で追うことのできる速度に基づいて一定のフレームレートの動画像で獲得を行うことを考えている。しかし現時点で最適なフレームレートを求められていないため、提案手法では一定方向に移動する対象に注目した場合のフレームレートを 3fps に固定している。この値は 1 秒間に 5 文字の速度で表示される文字列の情報を十分に獲得できるフレームレートに基づいている。



(a)絵画

(b)スクロールする文字列

図10 注目時の視線の移動

6 実験

6.1 実験方法

・決定したフレームレートの妥当性の評価

様々な視覚的に変化する対象に注目し、提案手法によって決定されたフレームレートと、各注目対象の視覚的な変化を捉えることのできる最適なフレームレートとの比較によって、提案手法の評価を行う。

用意した注目対象は、変化の穏やかな音楽プロモーションビデオ5本、TV映像(バラエティ番組、スポーツニュース、サッカー映像、映画、CM)の各種類を5本である。

ただし、各注目対象の視覚的な変化を捉えることのできる最適なフレームレートは、次のように求める。それぞれの対象について1fpsずつフレームレートを変化させて、1~10fpsの10個の動画像を撮影する。その後、対象の変化が十分に確認できるフレームレートを3名の被験者に決定してもらい、3名の平均値を視覚的な変化を捉えることのできる最適フレームレートとする。

・静止対象、及び一定方向へ移動する対象への注目の認識の評価

提案手法では、静止対象に注目した場合は静止画像、一定方向へ移動する対象に注目した場合は一定のフレームレートの動画像として獲得するため、それぞれの対象に注目し、静止対象かあるいは一定方向へ移動する対象に注目していると判断された割合によって、提案手法の評価を行う。

注目対象は、静止対象として、風景写真、絵画画像、人物写真の3種類の画像をそれぞれ10枚、一定方向へ移動する対象として、左方向、上方向に移動する文字列を用意し、それぞれの文字列の移動速度を1秒間に表示する文字数が3文字、4文字、5文字の3段階に変化させたものを用いた。なお、移動する文字列については、それぞれの場合について10回ずつ注目を行った。

6.2 実験結果及び考察

表1に視覚的に変化する各種類の対象に注目したときの最適フレームレートと提案手法で決定したフレームレートを示す。ただし、提案手法で決定したフレームレートは、各種類ごとに、それぞれの対象に注目して決定したフレームレートの平均値を示す。また、表2には静止対象に注目したときの結果、表3にはスクロールする文字列に注目した場合の認識率をそれぞれ示す。

表1より、提案手法によって決定した動画像のフレームレートがあらかじめ求めた最適フレームレートとほぼ等しくなる結果となった。しかし、映画については、提案手法で決定したフレームレートが最適フレームレートより低くなった。この原因として、映像中の変化の激しい部分が、視界画像中の有効視野領域において、小さな領域であったため、視覚的な変化を捉えることができなかつたことが挙げられる。

表2、3より、静止対象及び一定方向に移動する対象に注目した場合は、どちらの場合についても高い認識率を得ることができたが、一定方向に移動する対象については、左方向(水平方向)に移動する文字列の認識率が上方向(垂直方向)に移動する文字列の認識率よりもやや低くなる結果となった。この原因として、眼球を撮影するカメラが、眼球のやや下方に位置しているため、瞳孔の動きを検出する際、垂直方向に比べ水平方向の移動について、誤差が大きくなることが理由として挙げられる。

	最適FR [fps]	提案手法 [fps]
音楽PV(5本)	3	2.6
スポーツニュース(5本)	6	6.2
サッカー映像(5本)	6	5.7
バラエティ番組(5本)	5	5.0
映画(5本)	7	5.0
CM(5本)	5	5.2

表1 実験結果(視覚的に変化する対象)

	認識率[k/l]
風景写真(10枚)	1.0
絵画画像(10枚)	0.9
人物写真(10枚)	0.8

k: 静止対象に注目したと認識した回数

l: 注目した対象の数(10枚)

表2 実験結果(静止対象)

移動方向	速度[文字数/s]	認識率[n/m]
左方向	3	0.8
	4	0.9
	5	0.7
上方向	3	0.9
	4	1.0
	5	0.9

n: 一定方向へ移動する文字列への注目を認識した回数

m: 注目した回数(10回)

表3 実験結果(一定方向に移動する対象)

7 まとめと今後の課題

本研究では、視線の位置における視界画像の変化を検出することによって、注目した対象の視覚的な変化を検出し、その変化に応じて静止画像、及び変化を十分に捉えるフレームレートの動画像で視覚情報を獲得する手法を提案した。これにより、獲得した視覚情報から注目対象の視覚的な変化の様子を確認することが可能となる。また変化の度合いに応じて獲得する動画像のフレームレートを変化させることによって、ストレージを無駄に浪費することを防ぐ事が可能となる。

現在ユーザが静止している状態を前提としており、歩きながら注目を行った場合など、首や体の動きを伴う場合には対応できていないため、そのような場合であっても本手法が適用できるように改善することが今後の課題として挙げられる。また、注目対象がユーザから遠く離れていたり、小さい場合に、対象の変化を捉えることが難しくなるため、処理する画像のサイズの拡大等により対処する必要がある。

謝辞

本研究を行うにあたり、有益なご討論、ご助言を頂いた広島大学工学部データベース工学研究室の皆様へ感謝致します。

参考文献

- [1] A. Yoshitaka, M. Yoshida, “Automated Construction of Real World-Oriented Database with Gaze Detection”, Proc. of IEEE International Conference on Multimedia and Expo., 2001.
- [2] 池田光男, “眼はなにを見ているか”, 平凡社, 1998.
- [3] T. Jebara, B. Schiele, N. Oliver, A. Pentland, “DyPERS: Dynamic Personal Enhanced Reality System”, M.I.T. Media Lab. Perceptual Computing Section

Technical Report, No. 468, 1998.

- [4] T. Kawamura, Y. Kono, M. Kidode, “Wearable Interfaces for a Video Diary: towards Memory Retrieval, Exchange, and Transportation”, Proc. 6th International Symposium on Wearable Computers, 2002.
- [5] Y. Nakamura, J. Ohde, Y. Ohta, “Structuring Personal Activity Records based on Attention –Analyzing Videos from Head-mounted Camera”, Proc. 15th International Conference on Pattern Recognition, 2000.