社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2007－ＡＶＭ－59（18）
2007／12／14

視線情報に基づく映像要約手法の検討

応和 大輔† 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, ††{wataru,tominaga}@waseda.jp

あらまし 近年，VOD サービスの普及や PC の高性能化に伴い，デジタルコンテンツに触れる機会が増えている．膨大なマルチメディアコンテンツの効率的な利用と管理が求められる中，映像の内容を短時間で把握できる映像要約技術が注目されている．しかし，従来の多くの要約手法では映像や音声の特徴に依存しており，視聴するユーザの主観が考慮されているものは少ない．本研究ではユーザの視線情報を用いることで，ユーザの主観を反映した映像要約の手法を検討している．

キーワード 映像要約，早送り，瞳孔，注視点，テレビジョン番組

Video Summarization based on User's View Information

Daisuke OWA†, 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, ††{wataru,tominaga}@waseda.jp

Abstract With the recent growth of VOD service and advanced PC, we are increasingly watching digital content. Among need of efficient use and management of enormous multimedia content, we are paying a lot of attention to video summarization to understand content in short time. However, existing methods depend on only parameters of image or sound, and almost ignore user's subjectivty. In this paper, new method of video summarization which reflect user's subjectivity with user's view infomation is proposed and described.

Key words video summarization, fast-forward, pupil, gazing point, television program

1. はじめに

近年，地上デジタル放送の開始を皮切りに，ワンセグ放送や VOD サービスの普及が進み，ディジタルコンテンツは身近なものとなった．加え，PC の性能の飛躍的な向上や，大容量のストレージや高速なネットワーク回線などのインフラサービスが比較的低価格で一般にも提供されるようになったことで，さらなる拍車がかかっている．それに伴い，我々の視聴スタイルは時系列から解放され，従来のながら視聴から目的視聴にシフトしつつある．ホームサーバ等に蓄積された大量の映像コンテンツの効率的な管理と利用が求められる中，短時間で映像コンテンツの内容を把握できる映像要約技術が注目されている．

映像要約技術は主にニュース番組やスポーツ中継のような，知識のモデル化および内容の構造化が比較的容易なジャンルの映像を対象に研究が行われてきた．現在は様々な映像のジャンルに対し，映像特有のドメイン知識に基づく高度な内容解析による要約手法が数多く提案され，中には映像の種類を限定しない汎用的な手法も登場している．

しかし，これらの手法によって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，視聴するユーザにとって理解でき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観という要素を無視することは出来ず，また，前述のような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を反映させた映像要約の手法を検討する．

2. 研究背景

2.1 従来手法

映像要約技術はこれまでにも数多くの研究が行われており，その多くは映像の有するシンタックスおよびセマンティクスへのアプローチに大別される．図 1 にその代表的なフローを示す．

要約映像は映像をシーンチェンジや会話の切れ目で分割し，

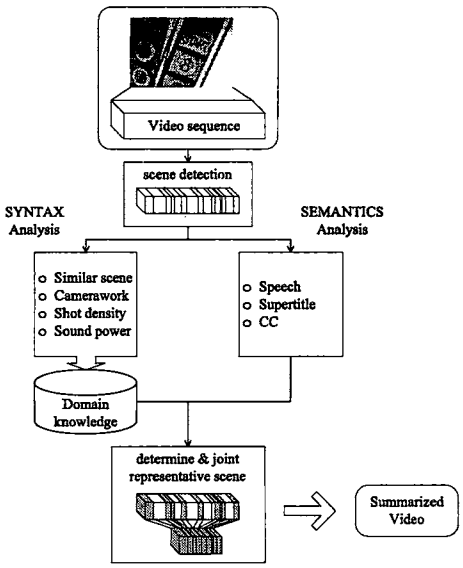

図 1　従来手法の代表的なフロー

重要度の高いシーンで再構成することで作成される．重要度の決定は主に類似シーンやカメラワークなどのローレベルな特徴を映像特有のドメイン知識と関連付け，特定のイベントを検出する手法と，発話内容や字幕，クローズドキャプションなどのハイレベルな情報を用いて映像の意味解析を行う手法に分けられる．

CMU の Informedia プロジェクトにおいては，画像，音声，CC(注1)の協調処理による映像スキミングが提案され，TF-IDF 法(注2)を用いて音声の書き下しである CC から重要なキーワードを抽出し，時間軸上で対応する映像および音声ストリームからカメラワーク（ズーム），顔やテキストの出現，音量の大きさなどの特徴を元に重要なシーンを選択，連結することで要約映像を作成している [1]．時間要約率が最大で 1/20 の要約映像の作成の報告がされている．MoCA プロジェクトにおいては，対象を映画に限定し，主要人物や重要なオブジェクトのクローズアップ，発砲音や爆発音などの特徴から特定のイベントを検出し，タイトルや主役による会話，アクションなどの重要なシーンをつなぎ合わせることで予告編的な要約映像を作成している [2]．

また，映像の内容解析といった高度な情報抽出処理は行わず，映像の持つ普遍的な特徴量やその変動などを考慮することによって，映像の種類に依存しない汎用的な要約手法も研究されている．分割したショットから動き情報や色情報などの映像特性および音声のパワーレベルといった音声特性を解析し，映画の文法 [3] といった映像制作の経験則から特徴パターンを検出し，代表的なシーンを抽出する手法が存在する [4]．この手法では，広く一般的に流通している MPEG の符号化データを用いることで，自動的かつ低コストで要約映像を抽出することができる．

しかし，要約映像の精度とアルゴリズムの汎用性はトレードオフの関係にあり，映像要約技術は映像の種類を限定した手法が主流であるといえる．

また，上述の要約手法は映像から特定のイベントを検出し，重要度の高いシーンで再構成する手法であるが，人間の映像再生速度の認知限界などを考慮し，比較的理解しやすいシーンを早送りにする手法も存在する [5]．早送りの手法は映像の文脈的構造をできるだけ損なわず要約することができるが，時間的な要約率は低い．

2.2　生理指標を用いた関連研究

一方，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．生理指標とは脳波や脈拍，呼吸といった脳や神経の機能を人間の心理活動と関連付けることで，今まで困難とされてきた人間の内面の客観的な観測を可能とするものである．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．しかし，万能というわけではなく，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の編集に活かす研究もすでに行われている．

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 成分の高い区間および低い区間をそれぞれつなぎ合わせる映像要約手法を提案している [6]．LF 値の高い区間をつなげた要約映像は心的負荷が少なく，理解しやすい映像になるという結果が得られている．

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 [7]．ユーザの状態をリアルタイムにフィードバックすることでユーザの興奮を効果的に誘起する映像や音声の作成が可能であるとしている．

また，中村らは固定カメラで撮影された HD サイズの映像の中から，ユーザの視線をもとに DV サイズの映像を切り出すことで，仮想的なカメラワークを加えた映像を作成している [8]．眼球運動を用いることでアマチュアでも容易にユーザ好みの映像を作成することができる．

3.　眼球運動

2.2 節でも述べたとおり，生理指標を用いた人間の心理活動の計測方法は数多く存在するが，中でも瞳孔径や視線の観測は比較的容易であることから，ハードおよびソフト両面から計測のための研究が進み，現在では国内外を問わず高性能な計測機器が開発されている．そこで本稿ではユーザの主観を抽出する方法として眼球運動を採用する．本節では眼球運動の特徴と生理指標としての眼球運動から人間の興味を検出する方法を示す．

3.1　特　　徴

眼球運動は概して跳躍運動，随従運動，固視微動の 3 つに分類される [9]．跳躍運動は視対象を変更する際に発生する眼球運

(注1)：Closed Caption
(注2)：Term Frequency Inverse Document Frequency

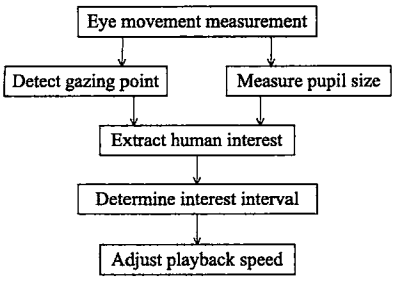

図 2　提案手法のブロック図

動で，サッケードとも呼ばれ，その速度は最大で 600[deg/sec] にもなる．随従運動は物体をゆっくり追尾する際に発生する眼球運動で，跳躍運動が高速なのに対し，随従運動は比較的低速で，その速度は最大でも 45[deg/sec] 程である．

以上の 2 つは対象物を追跡する機能であるが，対し固視微動は対象物を注視する機能で，一点を注視する際に発生する細やかな眼球運動である．固視微動はさらにその周波数や振幅，頻度によってフリック，トレモア，ドリフトの 3 つに分類される．

フリックは対象物の像が中心窩（ちゅうしんか）(注3)から大きく外れた際に，無意識に中心窩に像を結ぶために中心窩に向けて発生する．トレモアは対象物を中心窩上の一点に正確にとらえるのではなく，常に探索を行うために発生する．ドリフトは対象物の像を一定時間以上継続して中心窩に結ぶことができないため，像の移動として発生する．

眼球運動は追跡と注視が交互に発生する運動であり，視覚からの情報の大半はサッケードの間に得ているといえる．また，注視の際に発生する固視微動を人為的に止めてしまうと，視覚は物体を知覚できなくなってしまうことが知られている．

3.2　生理指標としての眼球運動

3.2.1　瞳　孔　径

人間は瞳孔の大きさを変化させることで外界から得る光量を調節している．しかし，Hess は対象物への興味度が高い場合には瞳孔が大きくなり，興味度が低い場合には瞳孔が小さくなることから，対象物への興味の度合いによっても瞳孔径が変化することを発見した [10]．これは，人間が本能的に興味度の高い対象物からより多くの情報を得ようとするためであると考えられている．瞳孔の大きさと人間の興味が比例関係にあることから，瞳孔径を人間の興味の指標とすることが可能であると考えられる．

3.2.2　注　視　点

人間は興味のある対象物を注視するという傾向から，山田らによって注視点の存在が定義された [11]．山田らは注視の機能である固視微動だけでなく，比較的低速な随従運動が発生する際にも対象物を認識できるとし，この 2 つの運動の発生によって視線が停留している状態を注視点として，人間が注視していると判断できるとしている．また，静止画だけではなく動画像においても注視の傾向は存在し，歌舞伎などの映像視聴時に注視パターンの個人差があることが確認されている [12]．映像視聴時における注視にも個人差があることから，注視点を用いて人間の興味を検出することが可能であると考えられる．

(注3)：網膜の中央に位置し，視野角 5[deg] 程度の解像度の最も高い部位

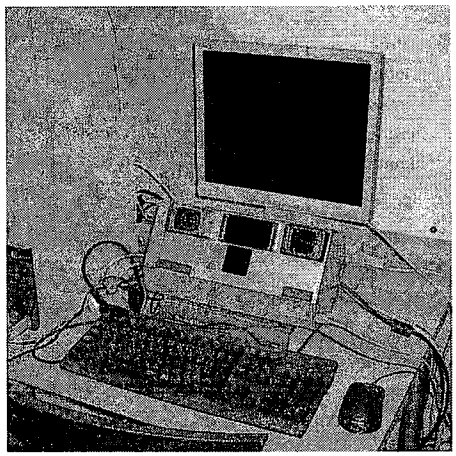

図 3　眼球運動計測システム外観

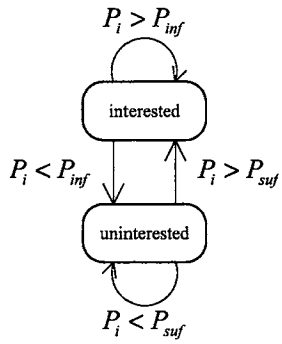

図 4　瞳孔径による興味状態の遷移図

4.　提 案 手 法

本節では生理指標を用いて眼球運動から人間の興味を抽出し，要約映像に反映させる手法として試作したプロトタイプシステムの説明をする．図 2 に本システムのブロック図を示す．本システムでは映像視聴時の瞳孔径と視点の位置を計測し，その計測結果から重要でないと判断されたシーンを早送りすることで要約映像を作成する．

また，本システムで用いた眼球運動計測システムの外観を図 3 に示す．本計測システムは検出器，刺激提示用 PC，制御用 PC からなり，角膜反射法(注4)と暗瞳孔法(注5)を用いて瞳孔および視点の位置を検出している [13]．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．

(注4)：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を利用して視線を検出する方法

(注5)：瞳孔と虹彩の輝度差から瞳孔を検出する方法

表 1　再生速度の割り当て

Pupil size	Gazing point	Playback speed
interested	interested	1-time
interested	uninterested	5-time
uninterested	interested	5-time
uninterested	uninterested	10-time

表 2　要　約　率

Test sequence	summarization ratio[%]	
	subject1	subject2
drama	22.70	25.36
variety1	19.55	16.40
anime	17.58	21.64
variety2	17.70	19.17
music	19.28	19.66

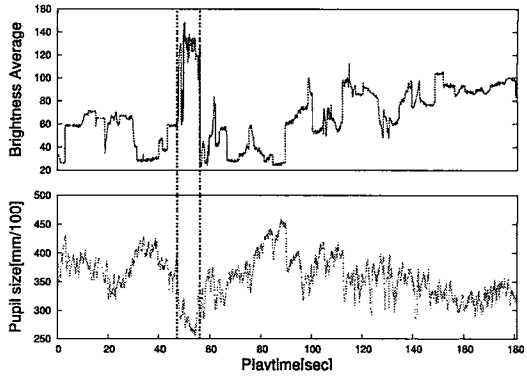

図 5　映像全体の平均輝度値と視聴時の瞳孔径

4.1　アルゴリズム

まず，映像視聴時の眼球運動の計測結動果から得られる瞳孔径および注視点に基づき，映像内におけるユーザの興味のある区間（以下，興味区間）およびない区間（以下，非興味区間）の検出を行う．

瞳孔径　計測結果から瞳孔径または視点の位置の計測に失敗した値をエラー値として除去し，エラー値を取り除いた計測結果から瞳孔径の平均値 P_{ave} を算出す動る．人間の瞳孔は対象物への興味度によってその径を ±5%〜10%増減させる．そこで，P_{ave} を平常時の瞳孔径とし，平均値の ±5%の値を興味の有無の閾値としてそれぞれ P_{sup} および P_{inf} と設定する．

$$P_{sup} = P_{ave} \times 1.05 \quad (1)$$

$$P_{inf} = P_{ave} \times 0.95 \quad (2)$$

瞳孔径の変動は固視微動による微弱な変動のような高周波のノイズを含むため，あるサンプル i において，サンプル i と前後の 2 サンプルずつ，計 5 サンプルにおける瞳孔径の平均値をサンプル i における瞳孔径 P_i とするスムージング処理を行った．

$$P_i = (P_{i-2} + P_{i-1} + P_i + P_{i+1} + P_{i+2}) \div 5 \quad (3)$$

スムージング処理を施した瞳孔径の計測結果を用いて，興味区間において P_{inf} を下回った区間を非興味区間，非興味区間において P_{sup} に上回った区間を興味区間と設定した．図 4 に瞳孔径による興味状態の遷移図を示す．

注視点　注視点は視点が半径 5[deg] の範囲内に停留している状態が 150[ms] 以上続いている状態と定義されている [11]．ここでは一般的に用いられる手法として，視点の移動速度が 5[deg/sec] 以下の状態が 150[ms] 以上続いている状態を注視点とする [14]．ここで，移動速度を視角からモニタ上の絶対位置に変換すると，計測システムのサンプリングレートは 60[Hz]，ユーザの顔とモニタとの距離は 0.8[m] であるので，モニタ上での視点の移動速度は約 300[pixel/sec] となる．さらにこれをサンプル単位で考えると，サンプル間の移動距離が 5[pixel] 以下の状態が 10 サンプル以上続いている状態が注視点となる．そこで，映像内において注視点の区間を興味区間，それ以外を非興味区間と設定した．

なお，作成された要約映像の短時間における頻繁な再生速度の変化を避けるため，区間長に閾値を設け，閾値以下の長さの区間を誤差として無視した．ここでは，瞳孔径における興味区間および非興味区間の区間長閾値を 1000[ms]，注視点における非興味区間の区間長閾値を 300[ms] とした．

両パラメータで検出された興味区間に用いて，その組み合わせによって再生速度を決定する．表 1 にパラメータ毎の興味区間の組み合わせによる再生速度の割り当てを示す．なお，本システムでは等倍，5 倍，10 倍の 3 段階で要約映像を作成した．

4.2　考　　察

本システムを用いて，被験者 2 名に対しテストシーケンス drama，variety1，anime，variety2，music の 5 件を提示した．上述の映像を視聴した際の眼球運動の計測結果から作成した要約映像の要約率を表 2 に示す．

本システムではユーザの眼球運動から興味が低いと判定された部分を早送りしているため，被験者毎の映像に対する興味の度合いの違いが要約映像の要約率に表れていると考えられる．しかし，主観の個人差が要約率に反映される反面，興味が高いと判定されるほど早送りは行われないため，要約映像の長さもユーザの興味の度合いに依存してしまい，予測が困難な一面も持つ．そのため，決まった時間内で映像を視聴する際には不向きであると考えられる．

また，本システムでは人間の興味というべきものを瞳孔径から検出したが，瞳孔径の変化は興味をはじめとする心理活動という内的要因による影響と外界からの光刺激という外的要因による影響が合わさった結果である．そのため，後者の外的要因に含まれる画面の明るさによる瞳孔径への影響も考慮しなければならない．図 5 に anime の画面全体の平均輝度値とある被験者の anime 視聴時の瞳孔径の変化を示す．映像全体の平均輝度値が急激に増加する 50[sec] 付近で，瞳孔径が大幅に減少しているのが確認できる．これは強い光刺激によって瞳孔径が急激に変化する対光反射のためであると考えられる．瞳孔径から人間の興味を抽出するには映像の輝度変化による瞳孔径への影響をノイズとして除去する必要がある．浅野らはこれらの問題に対してニューラルネットワークを用いて映像の輝度値から瞳

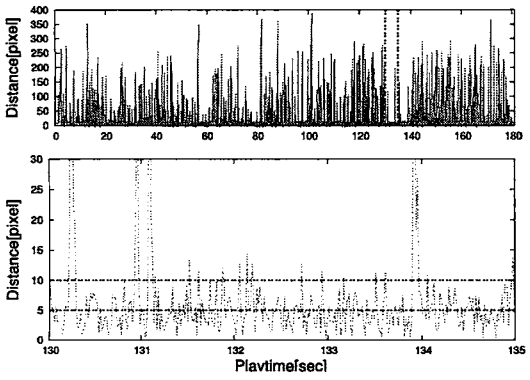

図 6　視聴時の視線の移動量

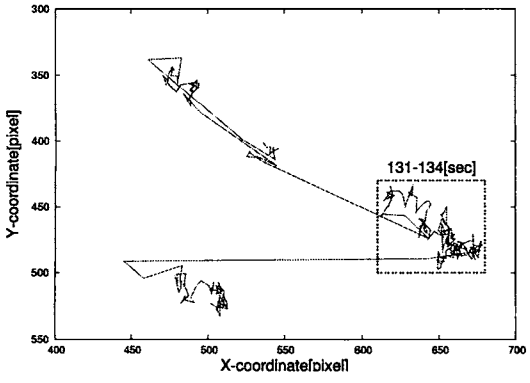

図 7　視聴時の視線の動き

孔径の値を補正する手法を提案している [15]. 浅野らの手法は単に瞳孔径を補正するだけでなく，提案するモデルが人間の生理学的な機構を模しているとしているため，有用であると考えられる.

図 6 にある被験者の anime 視聴時の視点の移動量を示す. 上段は映像全体における移動量，下段は 130~135[sec] における移動量である. 3.1 節でも述べたとおり，眼球運動は追跡と注視の 2 つの機能で構成されていることから，移動量の多い部分ではサッケードが，少ない部分では固視微動が発生していると考えられる. 本システムでは注視点の定義を視点の移動量に置き換え，その閾値を 5[pixel/sample] とした. サッケードの発生してない 131~134[sec] の区間に注目すると，移動量が閾値の 5[pixel] 付近を前後しているのが観察できる. また，注視点は円形の領域であるため，注視点内で想定される最大の移動量は円の直径であり，注視点の半径が 5[pixel] の場合は 10[pixel] となる. しかし，注視の最中でもその最大移動量を上回る場合も存在し，映像内における注視対象の動きも考慮すると，それが注視を外れているのかどうか厳密な判断を下すことは困難であると考えられる. 図 7 に同被験者の anime 視聴時の 130~135[sec] における視点の動きを示す. 図はモニタ全体ではなく，一部を拡大している. 131~134[sec] における視点の分布に注目すると，視点が集中しており，注視していると考えられる. 注視点の閾値を一意に決定することは難しいと考えられるため，一定以上の区間長を持つサッケード以外の部分を注視と判定し，その停留時間や停留頻度を参考にすることも可能であると考えられる.

5. むすび

本稿では人間の主観を抽出する方法として眼球運動を用いる手法を説明し，眼球運動をもとに要約映像を作成する手法とそのプロトタイプを提案した. 本システムでは映像視聴時の眼球運動をもとにユーザにとって重要でないと判断される部分を早送りすることで，ユーザの主観を反映させた要約映像を作成した.

本稿では映像視聴時の眼球運動の計測結果から要約映像を作成したが，生理指標を用いることで人間の情動反応を実時間で観測できるため，実時間で得られる計測結果をもとに映像の再生速度を変更することで，映像を視聴しながらにして要約映像を作成することも可能である. 人間の状態は個人間だけではなく，時間軸においても一意ではないため，嗜好の個人差だけでなく，個人の状態も加味することが可能であると考えられる. 今後，眼球運動を用いた動的な映像要約の検討と評価を行う.

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] M.A. Smith, T. Kanade, "Video skimming and characterization through the combination of image and language understanding, "Proc. ICCV1998, pp. 61-70, Jan 1998.

[2] R. Lienhart, S. Pfeiffer, W. Effelsberg, "Video abstracting, "Comm. ACM, Vol. 40, pp. 55-62, Dec 1997.

[3] Daniel Arijon, "Grammar of the Film Language, "Folcal Press.

[4] 菅野勝，中島康之，柳原広昌，"映像の特徴に応じた AV データからの自動要約抽出方式に関する検討，" 電子情報通信学会技術研究報告，Vol. 101, No. 494, pp. 25-30, Dec 2001.

[5] Kadir A. Peker, Ajay Divakaran, "Adaptive fast playback-based video skimming using a compressed-domain visual complexity measure, "Proc. ICME2004, Vol. 3, pp. 2055-2058, Jun 2004.

[6] 豊沢聡，河合隆史，"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.

[7] 杉田典大，田中明，阿部健一，吉沢誠，山家智之，仁田新一，"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.

[8] 中村亮太，井上亮文，市村哲，岡田謙一，"個人ビデオのための眼球運動を利用したデジタルカメラワーク，" 情報処理学会論文誌，Vol. 48, No. 1, pp. 163-170, Jun 2007.

[9] 大野健彦，"視線から何がわかるか - 視線測定に基づく高次認知処理の解明，" 認知科学，Vol. 9, No. 4, pp. 565-576, Dec 2002.

[10] E.H. Hess, "Attitude and pupil size, "Scientific American, Vol. 212, pp. 46-54, Apr 1965.

[11] 山田光穂，福田忠彦，"画像における注視点の定義と画像解析への応用，" 電子情報通信学会論文誌，Vol.J69-D, No. 9, pp. 1335-1342, Sep 1986.

[12] 松下清子，"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.

[13] "ナックイメージテクノロジー，"http://www.eyemark.jp/

lineup/EMR-AT/EMR-AT.html
[14] 福田忠彦, "生体情報システム論, " 産業図書.
[15] 浅野樹美, 安池一貫, 中山実, 清水康敬, "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.