First Person Visionにおける Inside-Outカメラ

関連動画
公開中!

正会員 藤吉 弘 亘[†], 木村 誠^{††}

1. ま え が き

近年、高性能カメラが小型化されたことにより、人の頭部や胸部等にカメラを装着して、装着者が見ているであろう視界に近い観測データから、装着者の行動を理解したり周囲の状況を認識しようとする研究が行われている。このような1人称視点を利用した画像解析、認識技術に基づく利用者への行動支援の枠組みをFirst Person Visionと呼ぶ。IEEEの国際会議Computer Vision and Pattern Recognition (CVPR) 2009にて、Workshop on Egocentric Vision¹⁾が開催され、First Person Visionの研究分野^{5) 14)}が注目を集めている*。

First Person Visionの研究では、人の頭部や胸部等にカメラを装着し、人の視点に近い位置から装着者が見ているであろう視界を観測する。この観測データを基にして、利用者の行動認識やライフログデータの収集、蓄積されたデータから解析や予測をして、利用者の行動の支援や生活の補助をすることが目的である。しかし、人の目は視界の一部に視線を向けて視対象を認識するため、視線情報を持たない観測データだけでは利用者の意図に沿う支援または補助をするには限界がある。この問題を解決するために、人の視線を計測する視線計測装置と、視界を観測するカメラを組合せた計測装置として、Inside-Outカメラが提案されている。

Inside-Outカメラは、図1に示すような、人の視点に近い位置から人の視界を撮影するSceneカメラと、人の眼球を撮影して視線計測等を行うためのEyeカメラにより構成されている。本稿では、First Person Visionに用いられるInside-OutカメラとInside-Outカメラの利用例について述べる。

図1 カメラとSceneカメラ

て述べる。

2. さまざまなInside-Outカメラ

Inside-Outカメラは、EyeカメラとSceneカメラの構成によって使用目的が変わり、図2に示すように、大きく非ゴーグル型とゴーグル型に分類できる。非ゴーグル型は、主にメガネ等の装飾品にカメラを搭載したタイプで、軽量で装着者の負担は少ないが、カメラを小型にする必要があるため、パフォーマンスは低くなる。ゴーグル型は、頭部に装置を固定し、複数のカメラを搭載することのできるタイプである。

そのため、複数のカメラの観測データを基に、利用者の視界や視線の情報を密に計測可能である。本稿では、SceneカメラとEyeカメラをそれぞれ1台搭載したInside-Outカメラを単眼型と呼び、Eyeカメラの代わりに赤外センサを搭載したタイプをセンサ型と呼ぶ。また、複数のSceneカメラとEyeカメラにより構成されるInside-Outカメラを複眼型と呼ぶ。

本章では、各Inside-Outカメラにとその応用について紹介する。

2.1 単眼型

単眼型のInside-Outカメラは、EyeカメラとSceneカメラがそれぞれ1台搭載されているタイプのカメラである。

* 2012年に第2回目のWorkshop on Egocentric Visionが開催予定。

† 中部大学 工学部 情報工学科

†† 株式会社サムスン横浜研究所

"Technology Frontier of Augmented Reality (4): Inside-Out Camera in First Person Vision" by Hironobu Fujiyoshi (Department of Computer Science, Chubu University, Aichi) and Makoto Kimura (Samsung Yokohama Research Institute Co., Ltd., Yokohama)

		← Inside-Outカメラ →				
		非ゴーグル型				ゴーグル型
デバイス
カメラ数		なし 1台	なし 1台	1台 1台	1台(センサ) 1台	2台 2台
視線推定		視線情報なし	頭部運動のみ	片目の視線(粗) 粗	片目の視線(密)	両目の視線(密) 密
提案手法			Badinoら ²⁾	金出ら ⁵⁾ Landら ^{6, 7)} Pelzら ¹⁰⁾ Sailerら ¹¹⁾	石黒ら ⁴⁾	Shneiderら ¹²⁾ 清水ら ¹⁵⁾ 西山ら ¹⁶⁾ 折笠ら ¹⁷⁾
応用先				3次元マッピング	医療	
				ユーザインタフェース		
				ロボットアイ		
				Augmented Reality (AR)		
				3D表示		
				車載カメラ		

図2 First Person VisionとInside-Outカメラ

図3 単眼型のInside-Outカメラ

1999年にLandらは、図3に示す単眼型のInside-Outカメラを用いて人の視線を計測し、人がお茶を淹れる動作において、無意識的に共通の視線パターンが存在することを発見している⁷⁾。このような「視線の動きを調べる」研究は、1967年にYabusが「写真(の中のコンテンツ)を注視する」という被験者実験で視線の動きを調べたことが起源とされ、2001年にPelzらは、人の複雑な動作として洗面台で手を洗うという仕草の観測結果から、シーンの一部で視線

図4 石黒らのInside-Outカメラ(文献4)より抜粋

が先の動作場所へ動くという結果を報告している¹⁰⁾。

2005年には、Sailerらは知恵の輪のような簡単なタスクで知覚運動の学習をさせると、習熟に伴って視線の動きが変化することを報告している¹¹⁾。2006年にLandらは、ドライバが車を運転するシーケンスから、ハンドルの操作時に視線と体の動作に相関があることを報告している⁶⁾。

近年では、金出らが、生活の質の向上(Quality of Life:QoL)を目的としたFirst Person Vision Wearable Deviceとして、メガネにEye, Sceneカメラを1台ずつ搭載させた単眼型のInside-Outカメラを提案し¹⁾、Inside-Outカメラを用いた視線解析以外の利用法について言及している。

2.2 センサ型

石黒らは、図4に示すように、単眼型のInside-OutカメラにEyeカメラではなく、光センサであるフォトトランジスタを搭載したセンサ型のInside-Outカメラを提案している⁴⁾。センサを用いることで、一般のカメラでは捉えるこ

図5 複眼型のInside-Outカメラ(文献12)15)より抜粋)

とができない高速な眼球運動や瞬きの検出を実現している。センサ型は、カメラよりも小型化しやすく、低い計算コストで処理できる点が特徴であるが、環境光によるノイズの影響を受けやすい点を考慮する必要がある。文献4)において石黒らは、Inside-Outカメラを用いたライフログシステムを提案しており、人や物といった利用者に合った事前知識を登録しておくことで、注視領域内に存在する対象を高速に認識できるシステムとなっている。さらに、顔認識や文字認識を組合せることで、顔や文字などの認識対象の種類を拡張することも可能である。また、認識結果を蓄積することで、利用者の特徴に合わせた行動支援が可能となることから、データマイニングや物体認識への展開も考えられる。

2.3 複眼型

Schneiderらによって提案されたEyeSeeCam¹²⁾は、Eyeカメラ2台とSceneカメラ2台で構成されている。図5(a)に示すように、視界撮影用のカメラ1台は、人の視野全体を捉える広い画角を持ち、もう1台は、推定された注視点を追跡するPTZカメラとなっている。

EyeSeeCamは、両目の眼球映像と視界全体の映像に加えて、人の注視領域の高解像度な映像が撮影できるため、応用先は行動解析やライフログなど多岐にわたる。例として、熟練者の作業(医師の施術や看護師の機器確認)の注視点映像を撮影することで、ログデータを、経験の少ない新人や若手に対する教育目的に利用することができる。また、前庭動眼反射(VOR)実験では、頭部運動を打ち消すよう

な眼球運動を観測することで、健常者と異常者の診断にも利用されている。

著者の研究室では、図5(b)に示すような、ゴーグル型のフレーム上部にEyeカメラ2台と、下部に配置したSceneカメラ2台で構成されたInside-Outカメラを提案している¹⁵⁾。

このInside-Outカメラは、図5に示すように、眼球とEyeカメラの間で赤外光のみを反射するIRミラーと、シーンからの可視光をSceneカメラと眼球に分岐するハーフミラーが中央部に設置されているため、透明なカメラが存在するような光学的配置を実現している。著者らは、このInside-Outカメラを用いて、注視点の3次元位置を推定する手法を提案している¹⁵⁾。

3. Inside-Outカメラによる3次元注視点推定

本章では、著者らが提案しているInside-Outカメラにおける3次元注視点推定の手法について述べる。筆者らのInside-Outカメラは、他のInside-Outカメラとは異なり、Sceneカメラの位置が視線の位置とほぼ等しくなるよう配置されているため、シーン映像は人の見ている視界と同等に扱うことが可能である。そのため、カメラの位置ずれを考慮する必要がなく、眼球とEyeカメラ、Sceneカメラそれぞれの校正が容易に行える特徴がある。以下に、Inside-Outカメラによる3次元注視点推定の流れを示す。

(1) 視線ベクトル推定：瞳孔中心と角膜曲率中心から視線ベクトルを推定

図6 注視点の算出の処理

(2) 注視点の算出：注視対象と視線ベクトルの相関関係から注視点を算出

(3) 次元注視点の算出：Eyeカメラが2台ある場合、両眼の注視点を用いてステレオマッチングにより3次元の注視点が算出可能

3.1 視線ベクトルの推定

視線検出手法により、眼球に赤外光が反射して映るブルキニエ像から角膜曲率中心を算出する。この角膜曲率中心と瞳孔中心を結ぶ直線を視線ベクトルとする。視線ベクトルは、角膜曲率中心とブルキニエ像から推定可能である。ブルキニエ像から角膜曲率中心を算出し、瞳孔中心を推定する。そして、角膜曲率中心と瞳孔中心から視線ベクトルを推定する。

3.2 変換式を用いた注視点の算出

3.1節により求めた視線ベクトルを用いて、画像上の注視点を推定する。

ここでは、変換式の変換パラメータを求める処理と、変換式を用いた注視点を推定する処理の、二つに分けて述べる。その処理の流れを図6に示す。

まず、視線ベクトルと注視点の関係を変換式で表現するパラメータを推定するため、視線ベクトルを入力としたときの注視点を算出する。次に、算出した変換パラメータを用いて、視線ベクトルから注視点を推定する。そして、算出した視線ベクトルを各成分に分割し、注視点を算出する。

3.3 3次元注視点推定

左右のEyeカメラから算出される注視点を、ステレオ視の問題に置き換えることで、3次元注視点を推定する。瞳孔中心と注視点を結ぶ線を視線とし、左右それぞれの視線の延長線上で交わる点が3次元の注視点となる。しかし、図6に示すように、左右の視線は正確に交わらないため、著者らは、最適補正により、二つの注視点と視線が対応するように位置の補正を行い、3次元注視点位置を推定している。図7に、著者らによる3次元注視点推定の結果を示す。3次元の注視点を推定することで、従来は2次元で計測されていたScan Path¹³⁾のような視覚機能を、3次元

図7 3次元注視点の推定結果

的に計測・解析することができると考えられる。3次元の注視点の解析により、従来行われている視線計測や行動解析に適用することで、注視点に奥行き情報を含んだ解析結果が得ることができる。

4. 今後の展望

これまで述べたInside-Outカメラによる研究は、First Workshop on Egocentric Vision¹⁾で、金出らが提案しているように、Sceneカメラによるライフログ収集や、収集されたデータから建築物のマッピングなどの応用が考えられる。本章では、将来予測も含めた各分野における今後

の展望について述べる。

4.1 医療

2.3節で述べたEyeSeeCamは、元々医療研究用に開発されているため、さまざまな医療目的の応用が考えられている。例えば、多様な特性の色覚異常者を被験者として、自然環境中にさまざまな色のオブジェクトを配置し、どのような色に対してどのように視覚が反応するかを調べるような調査を行っている。このように、Inside-Outカメラでは、自然環境でのデータ取得が可能になるため、医療目的（診断、リハビリなど）の応用は今後も発展が期待できる。

4.2 3次元マッピング

Inside-Outカメラを装着した利用者が、歩行などの動作をすることで、多視点によるカメラ映像が取得できる。そのような無拘束の多視点映像から、特徴点抽出などの処理を経て3D形状を復元する研究として、Structure from Motion (SfM) が有名であるが、2011年に、人の運動モデルを考慮した3D形状復元手法をBadinoが提案した²⁾。

4.3 ユーザインタフェース

従来、視線計測システムに関する研究は数多く存在している。例えば、NAC社のEMR-AT Voxerのように、注視対象をディスプレイなどの平面に限定することで、非接触・無拘束な条件で、画面の中の注視点位置を特定するシステムがある。このような「ディスプレイ中の注視点の検出」というシステムは、研究や解析などを目的とした視点検出だけではなく、ユーザインタフェースとしての応用も期待されている。Inside-Outカメラを視線計測システムの拡張として考えると、Inside-Outカメラの応用によって、より一般的な条件下での視線ユーザインタフェースの実現が可能になると考えられる。

一方、ここ数年で多種多様な小型カメラ（例：メガネ型、耳掛け型、自転車のマウント用、ヘルメットへのマウント用など）が安価に流通するようになり、1人称視点からの映像の記録が容易になってきている。例えば、サーフィンやハングライダーなど、一般人が体験し得ない視点の映像が撮影できるため、YouTubeなどで1人称視点の動画が話題になることも多い。Inside-Outカメラを、このようなユーザ視点映像の拡張として考えると、2.3節で述べたEyeSeeCamの医療訓練のような熟練者の練習だけでなく、スポーツの練習などへの応用も考えられる。

4.4 ロボットアイ

EyeSeeCamの提案者であるSchneiderらは、人間の視線の動きを低レイテンシで再現できるロボットを作成し、Hannover Messe 2009において、EyeSeeCamでロボットの視線を遠隔制御するデモを行った。Schneiderらの研究では、このような視線制御を加えたロボットをアバターとして用いることで、遠隔コミュニケーションの質の向上が確認されている。また、人の視線の先をロボットが認識する「共同注視」という研究があり、多視点の監視システム

や、利用者の死角のアシストなどの応用分野が考えられている。ここにもInside-Outカメラは応用可能であると考えられる。

4.5 AR

従来、Virtual Reality (VR) やAugmented Reality (AR) において、ヘッドマウントディスプレイ (HMD) を用いたシステムは、研究、商用含めて数多く存在している。このようなHMD装着型のシステムにおいて、視線計測も含めたInside-Outカメラを融合することは、自然な流れであると予想される。

4.6 3D表示

ここ数年でメガネ型の3D-TVが多数発売されているが、それらの製品は、両眼視差のみで3Dを表現しているため、3D表示としてのさまざまな問題点が指摘されている。例えば、「運動視差がないため、頭部の移動に対して不自然になる」「実際の注視点の奥行き（表示面）と映像のフォーカスの不一致による眼球疲労」などが有名である。前者に対しては、Wiiリモコンを用いた頭部位置の観測や、iPhone、iPadの顔認識などを用いて、運動視差のみで3D的な表現を実現したシステムが提案されている。後者に対しては、注視点位置に合わせてフォーカスポケを調整して提示するシステムが提案されている¹⁷⁾。1人称システムを用いると、ディスプレイに対するユーザ視点位置の観測とディスプレイ内の注視点位置の推定を同時に行うことが可能になるため、3D表示の向上への応用が考えられる。

4.7 車載カメラ

近年、ドライブレコーダと呼ばれる車載カメラが普及している。これは、事故が起きた場合に、その直前の状況を確認することを主目的としたカメラである。一方、ドライバの居眠り検出や、眠気の予測¹⁶⁾などの研究は、過去に数多く行われている。これら「車の外の映像」、「ドライバの眼球の計測」を同時に行うカメラシステムは、ある種のInside-Outカメラと考えることができる。また、もしもInside-Outカメラの装着によって安全性を高めることが可能であれば、将来的に、ドライバ頭部へのInside-Outカメラの装着という展開も考えられるであろう。

5. む す び

本稿では、First Person Visionの研究に用いられるInside-Outカメラと各研究について紹介した。1人称カメラや視線計測の分野は、カメラやコンピュータの性能向上に伴い急速な発展を遂げており、実に多様な取組みが行われるようになってきている。近年、視線計測の分野では、マイクロサッカードを計測することで人の潜在意識を解析する研究が進められている^{3) 8) 9)}。マイクロサッカードは、人が視線を動かす場合に、無意識に興味の対象へ小刻みに向く眼球運動である。そのため、二者択一のような実験から、人の内面、潜在意識を解析する手がかりとなる眼球運動と

して期待されている。マイクロサッカードは、極めて微小(0.1~1deg)で高速(500fps)な眼球運動であり、今後は、マイクロサッカードを観測するために、Inside-Outカメラの高解像度化かつ高速化が期待される。

First Person Visionの研究は、さまざまな分野や目的に適用することが可能であり、今後、私たちの生活の質の向上(Quality of Life: QoL)に欠かせない技術として、いっそうの研究開発が期待される。

(2012年2月6日受付)

〔文 献〕

- 1) First Workshop on Egocentric Vision, in conjunction with CVPR2009 (2009) <http://www.seattle.intel-research.net/egovision09/>
- 2) H. Badino and T. Kanade: "A head-wearable short-baseline stereo system for the simultaneous estimation of structure and motion", In IAPR Conference on Machine Vision Application (MVA) (2011)
- 3) K. Inagaki, Y. Hirata and S. Usui: "A model-based theory on the signal transformation for microsaccade generation", Neural Netw, 24, pp.990-997 (Nov. 2011)
- 4) Y. Ishiguro, A. Mujibiyi, T. Miyaki and J. Rekimoto: "Aided eyes: Eye activity sensing for daily life", In Proceedings of the 1st Augmented Human International Conference, p.25 (2010)
- 5) T. Kanade: "First-person, inside-out vision" (2009)
- 6) M. Land: "Eye movements and the control of actions in everyday life", Progress in Retinal and Eye Research, 25, 3, pp.296-324 (2006)
- 7) M. Land, N. Mennie, J. Rusted, et al: "The roles of vision and eye movements in the control of activities of daily living", PERCEPTION-LONDON-, 28, 11, pp.1311-1328 (1999)
- 8) S. Martinez-Conde, S. Macknik, X. Troncoso and D. Hubel: "Microsaccades: A neurophysiological analysis", Trends in neurosciences, 32, 9, pp.463-475 (2009)
- 9) J. Otero-Millan, S. Macknik, A. Serra, R. Leigh and S. Martinez-Conde: "Triggering mechanisms in microsaccade and saccade generation: A novel proposal", Annals of the New York Academy of Sciences, 1233, 1, pp.107-116 (2011)
- 10) J. Pelz and R. Canosa: "Oculomotor behavior and perceptual strategies in complex tasks", Vision Research, 41, 25-26, pp.3587-3596 (2001)

- 11) U. Sailer, J. Flanagan and R. Johansson: "Eye-hand coordination during learning of a novel visuomotor task", The Journal of neuroscience, 25, 39, pp.8833-8842 (2005)
- 12) E. Schneider, T. Villgratner, J. Vockeroth, K. Bartl, S. Kohlbecher, S. Bardins, H. Ulbrich and T. Brandt: "Eyesecam: An eye movement-driven head camera for the examination of natural visual exploration", Annals of the New York Academy of Sciences, 1164, 1, pp.461-467 (2009)
- 13) A. Yarbus: "Eye movements and vision", Plenum press (1967)
- 14) 亀田能成, 藤吉弘亘, 羽下哲司, 山口晃一郎, 中村裕一: "First Person Visionへの挑戦 (First Person Visionのための認識・理解)", パターン認識・メディア理解研究会 (PRMU), 110, 28, pp.37-38 (2010)
- 15) 清水彰一, 西尾和見, 木村誠, 藤吉弘亘: "First Person VisionのためのInside-Outカメラの提案", 信学誌D, J94-D, 11, pp.1909-1918 (2011)
- 16) 西山潤平, 木下伸一, 平田豊: "前庭動眼反射による眠気予兆検出", 生体医工学, 48, 1, pp.1-10 (2010)
- 17) 折笠達郎, 鈴木孝明, 岡谷貴之, 出口光一郎: "注視反応ディスプレイによる両眼視差とボケの同時提示", コンピュータビジョンとイメージメディア (CVIM), 2011, 40, pp.1-8 (2011)

藤吉 弘亘 (ふじよし ひろのぶ) 1997年, 中部大学大学院博士後期課程修了。1997~2000年, 米カーネギーメロン大学ロボット工学研究所Postdoctoral Fellow。2000年, 中部大学講師, 2004年, 同大准教授を経て, 2010年より, 同大学教授。2006年, 米カーネギーメロン大学ロボット工学研究所客員研究員。計算機視覚, 動画像処理, パターン認識・理解の研究に従事。2005年度, ロボカップ研究賞。2009年度, 情報処理学会論文誌コンピュータビジョンとイメージメディア優秀論文賞。2009年度, 山下記念研究賞。2010年度, 画像センシングシンポジウム優秀学術賞。博士(工学)。正会員。

木村 誠 (きむら まこと) 2002年, 慶應義塾大学大学院理工学研究科博士課程修了。同年, 科学技術振興事業団CREST研究員。2006年, 産業技術総合研究所研究員を経て, 2010年より, (株)サムスン横浜研究所。コンピュータビジョン, 画像処理に関する研究開発に従事。2010年, ファナックFAロボット財団論文賞。2010年, 計測自動制御学会計測部門論文賞。博士(工学)。