

視線方向と身体動作の関係性を利用した間接的視線推定

代表研究者 山 添 大 丈 兵庫県立大学 大学院工学研究科 准教授

1 はじめに

視線情報には、人の注意や興味、意図などの様々な情報が含まれ、人の内部状態の理解に有用であるため、様々な分野において利用されている[1]。例えば、マーケティングや広告の分野においては、顧客の興味対象を推定する手段として視線が用いられている[2]。ヒューマン・コンピュータ/ロボット・インタラクションにおいては、適切なインタラクション実現のために視線情報を利用している[3]。また、視線からドライバの眠気の有無や不注意、認知負荷を検出・推定することで運転支援を目指した研究[4]や、自閉症や統合失調症の診断などの医療への応用を目指した研究[5]なども行われている。

このように、視線情報には様々な応用があるため、これまでに、多くの画像に基づく視線推定の研究が進められており、それらに基づく市販システムも多く存在する。これらの視線推定手法は、推定の仕組みによって大きく二種類、赤外光源に基づく手法とカメラ画像のみに基づく手法に分けられる。赤外光源に基づく手法では、その多くで瞳孔角膜反射法[6]を用いている。これは、近赤外の点光源(LED)により目を照射し、角膜で反射された光源像(プルキニエ像)と瞳孔の位置から視線方向を推定する手法である。推定誤差が1度未満と比較的高精度な視線推定を実現しているが、赤外光源によって目を照明し、その反射像を撮影する必要があるため、視線の検出範囲がカメラ(赤外光源)の近傍に限られる、プルキニエ像を検出するために目領域の高解像度画像が必要という問題が存在する。これに対し、カメラ画像のみに基づく手法では、計測対象人物の目領域や顔領域の画像が撮影できれば良いため、計測の仕組み上の距離の制約はない。近年では、深層学習を用いた手法も多く提案されているが[7-10]。ただし、画像のみに基づく手法でも、計測対象人物の目領域を観測できる位置にカメラを設置する必要がある。

カメラの設置位置によっても、視線推定手法は大きく二種類に分類でき、装着型と据置型に分けられる。装着型は、計測対象人物の眼前方にカメラを装着し視線方向を推定するもので、広範囲で視線が取得できる。移動する人物の視線方向を推定することも可能である。しかし、目前方にカメラが装着されているために、視野の妨げになるという問題が存在する。据置型では、計測対象人物が装置を装着する必要はないが、視線の取得が装置周辺に限られ、移動する人物の視線方向を推定することは難しい。このように、従来手法では、視野の妨げにならずに、移動する人物の視線方向の推定を行うことは難しかった。

これに対し本研究では、目領域を直接観測して視線方向を推定するのではなく、間接的に推定することを考える。提案手法の基本的なアイデアは、視線方向の変化によって他の身体部位の動作に変化が生じることを利用し、それら他の身体部位の動作の観測から、間接的に視線方向を取得するというものである。具体的には、図1に示すように、計測対象人物の胸部に広視野(魚眼)カメラを装着し、その画像から装着者の身体動作を取得し、それら身体動作の変化から視線方向を間接的に推定する。本研究期間では、この実現に向けて、装着カメラによる身体動作(頭部・腕動作)の推定と、視線方向の変化に伴う身体動作の変化のモデル化のための実験環境の構築を進めた。

2 関連研究

2-1 ウェアラブルカメラによる身体動作の推定

ウェアラブルカメラを用いて人の行動を推定・認識することを目指した多くの研究が行われている。その中には、本研究で考えるような身体動作の姿勢推定を目指した研究も存在し、例えば、Shiratoriらの研究[11]では、人体の各部位にカメラを装着し、周囲の環境を観測することで、間接的に人体各部位の姿勢推定を実現している。しかし、姿勢を推定したい部位全てにカメラやセンサを装着する必要があり、容易に利用できるシステムとはなっていない。

これに対し、広視野カメラを計測対象人物に装着し、装着者自身の動作を観測することによって、全身姿勢を推定する手法がいくつか提案されている[12,13]。Xuらは、帽子に下向きに装着した魚眼カメラを用いた全身の姿勢推定手法を提案している[12]。Tomeらは、Xuらの手法と同様に、ヘッドマウントディスプレイ

(HMD)に下向きに装着した魚眼カメラを用いた全身の姿勢推定手法を提案している[13].これらの手法では、帽子のつばやヘルメット、HMDなどにカメラを装着し、より前方の視点から全身を撮影することで、より歪みが少ない画像を取得している。これらは高精度な姿勢推定に寄与している。一方で、頭部にカメラを装着すると、カメラが計測対象人物の視野の妨げになるという、装着型の視線推定手法と同様の問題が生じる。

頭部にカメラを装着する必要がある手法として、我々は胸部に装着したカメラを用いた頭部姿勢推定や指差し方向推定などの研究を行ってきた[14, 15].胸部にカメラを装着することで、脚の姿勢推定は難しくなるが、装着者への負荷は低減するという利点がある。同様のアプローチとして、Hwangらも胸部に装着したカメラを利用して全身の姿勢推定を目指した手法を提案している[16].

2-2 視線方向変化に伴う身体動作変化

人は目を動かすことで、周辺環境の視覚情報を取得しており、より広い範囲の情報を取得する際には、目だけでなく頭部も動かしている。その際、目と頭部は協調的に運動するため、その関係性を調査した多くの研究が存在する。例えばStahlらは、実験環境と自然な環境の両方で、頭部方向と視線方向を計測し、眼球運動と頭部運動の振幅に関係性が存在することを報告している[17, 18].

さらに広い範囲を見る場合には、目・頭部に加えて胸部(胴体)の運動も生じており、これらの運動の関係性を分析した研究も多く存在する。例えばAnastasopoulosらは、目・頭部・胸部の運動を分析し、目・頭部・胸部の3つの相対姿勢が2自由度しか持たないことを報告している[19].また、歩行時には、視線を安定させるために頭部と胸部が協調的に運動することも知られている。そのため、直線歩行時や円の軌跡で歩行時など、様々な歩行条件での分析も行われている[20].

これらを踏まえ我々も、トレッドミル歩行時に意図的に注視方向を変化させたときに、全身の運動にどのような影響が生じるかを調査し、腕振りや頭部・胸部方向と視線方向に関係性があることを確認した[21, 22].このような関係性を利用し、間接的に視線方向の推定を目指すのが本研究の基本的なアイデアである。

3 視線方向と身体動作の関係性を利用した間接的視線推定

3-1 提案手法の流れ

提案手法の流れを図1に示す。これまでの研究[21, 22]より、視線方向の変化により、腕振り、頭部・胸部方向に影響が生じることが分かっており、これらの関係性を利用して目領域を直接観測することなく、間接的に視線推定を行う。まず胸部に装着したカメラの画像から腕の動作と頭部方向を推定する。次に、視線方向と身体動作の関係性に基づき、視線方向を推定する。

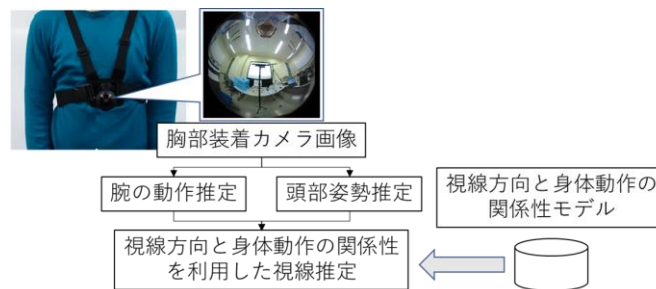


図1 提案手法の概要

3-2 胸部装着カメラによる腕姿勢推定

胸部装着カメラによる腕動作の推定手法としては、[15]で提案した手法をベースとする。この手法[15]では、図1上に示すように胸部に装着された魚眼カメラを用いて腕部位(手先・肘・肩)の画像中の2次元座標を抽出し、これらの座標から指差し方向を推定している。水平・垂直方向ともに平均7度程度で指差し方向を推定できることを確認している。

本研究では、指差し方向ではなく、腕部位の座標を直接利用するが、その際いくつかの問題が生じる。一つは、[15]では指差し動作の推定を対象としており、推定モデル構築のための学習データは指差し動作のみを含んだものとなっており、歩行中の腕振りといった指差し動作以外の腕の動作に対応できないことである。

もう一つは、腕部位の座標として推定されるのは画像中の 2 次元座標であり、3 次元姿勢としては推定されていないことである。

これらを解決するため、歩行中の腕振りなどの指差し動作以外の動作を含む画像データを取得し、それを用いて推定モデルを再学習することで、指差し動作以外にも対応した推定モデルを構築することを考える。その際、学習用データ数を増やすために、CG モデルを用いて様々な腕の動きを含む画像データを生成する方法についても準備を進めている。図 2 に生成画像データの例を示す。

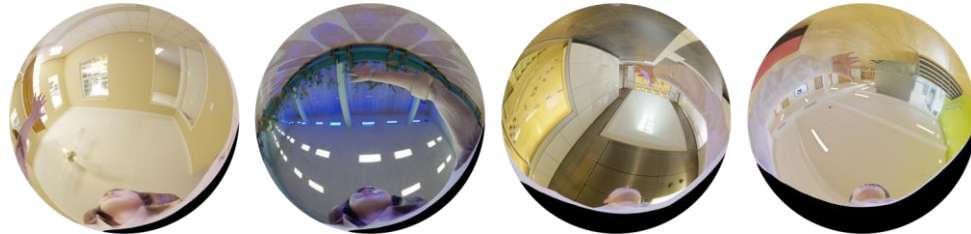


図 2 胸部装着カメラ画像の生成データの例

3-3 胸部装着カメラによる頭部方向推定と個人適応

次に、胸部装着カメラによる頭部方向の推定について説明する。視線方向との関係性において、頭部方向は特に視線方向との関連が高いため、個人差による影響を低減することが重要と考えた。そこで、個人に適応した頭部姿勢推定の枠組みを検討した。図 3 に頭部方向推定の流れを示す。今回の実装では、入力画像に対する処理としては、画像認識のための深層学習モデルの一つである VGG16[23]を用いることとし、VGG16 の後ろに全結合層 2 段を追加し、その出力と頭部形状パラメータを結合したのに対し、さらに全結合層 1 段を適用するネットワーク構成とした。このネットワーク構成は、個人適応した頭部姿勢推定手法[24]を参考にしたものであり、入力画像と頭部形状パラメータを入力として、頭部形状に応じた頭部方向を推定する。これにより、個人に適応した推定を実現している。

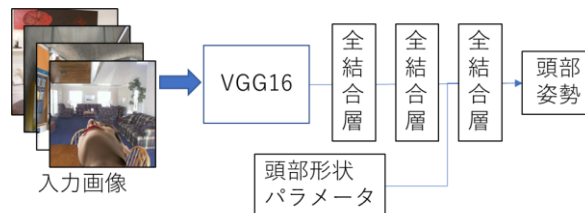


図 3 個人適応した頭部方向推定

3-4 視線方向の変化による人体姿勢の変化のモデル化

視線方向と身体動作の関係性のモデル化については、これまでの研究[21, 22]では、視線方向の変化によって身体動作（頭部・胸部・腕振り）に影響が生じ、両者間には線形の関係があることを確認していた。ただし、直線歩行時（トレッドミル歩行時）に限定であったこと、歩行周期ごとの特徴量（平均姿勢・振幅）に着目した分析となっていたことなど、本研究で想定する視線推定手法での利用に適したモデルとはなっていない。

そこで、自由に行動する被験者の全身の姿勢と視線方向を取得し、その関係性をモデル化するため、VR に基づく実験環境を構築した（図 4）。VR 環境中で視覚探索タスクを行ってもらい、その際の視線方向・全身の姿勢を記録する。VR 環境の構築には、視線計測機能付ヘッドマウントディスプレイ（HTC Vive Pro Eye）と VR 用トラック（HTC VIVE トラック 2018）を用いており、Unity（2019.4.10f1）、VR 用アセット（SteamVR、ViveSR）を用いて実装した。VR 用トラックは、全身の姿勢を取得するために利用し、VR 用トラック 8 個を左右の手首・前腕・足首と胸前方、腰後方に装着している。頭部姿勢は HMD に搭載されたセンサから取得できる。VR 用トラックから 3 次元位置・姿勢が取得できるため、これらのデータから逆運動学を計算する Unity 用アセット（Final IK）を利用して全身姿勢の推定を可能とした。



図 4 VR 実験環境

4 実験

4-1 胸部装着カメラによる頭部方向推定と個人適応

まず、3-3 節で説明した頭部方向推定モデルによる推定精度を評価するための実験を行った。推定モデルの学習のためのデータとしては、多くの被験者による学習データの取得は時間・手間がかかるため、3Dモデルを用いて生成した画像を用いた。画像データの生成には、まず顔姿勢推定・3次元形状復元手法の一つである文献[25]の手法を用いて顔形状の3次元復元を行う。次に、復元された顔の3次元モデルを様々な姿勢(水平方向 ± 30 度, 垂直方向 ± 10 度)で配置し、それを仮想の胸部装着カメラでレンダリングすることで、様々な頭部姿勢の画像列を生成した。頭部姿勢推定ネットワークに入力される頭部形状パラメータとしては、手法[25]から得られる40次元の顔形状パラメータを利用した。データ生成の元となる顔画像としては、AFLW2000-3D[25]に含まれる2000名分の画像を用いた。また様々な背景を含む画像とするため、屋内シーン推定のためのデータセット[26]を背景画像として利用した。図2に生成された画像列の例を示す。



図 5 頭部方向推定モデル学習用の生成データの例

これらの画像列を用いて頭部方向の推定モデルを構築した。生成画像のうち306621枚(371名分)を学習用、87885枚(135名分)を検証用、5000枚を評価用として用いた。学習用・検証用・評価用の分割にあたっては、同一人物の画像が含まれないようにした。画像サイズは150×150ピクセルとした。また、個人適応の有効性を評価するため、図6に示す推定モデルに加えて、ベースラインとして頭部形状パラメータを用いないモデル(VGG16の後、2層の全結合層のみ)についても同じ学習用画像データで学習し、精度を比較した。結果の一部を図8,9に示す。その結果、ベースライン手法では、平均誤差約1.27度で推定が可能だったのに対し、提案手法では、平均誤差約1.15度と、ベースライン手法と比較して、わずかではあるが誤差が小さく、個人適応により頭部方向の推定精度が向上することが確認できた。

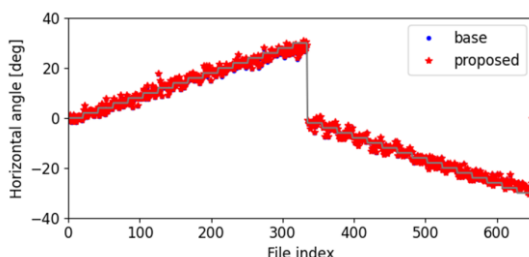


図 8 水平方向姿勢の比較

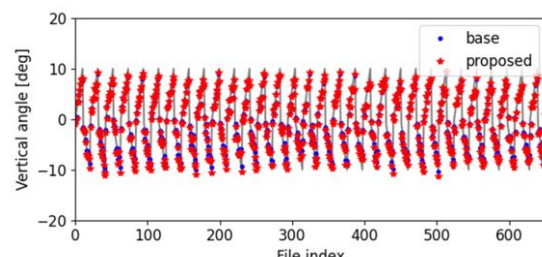


図 7 垂直方向姿勢の比較

文献[24]では、同様の構成のネットワークを用い、個人の形状パラメータも入力することで、個人適応した顔姿勢推定を可能としており、顔の撮影方向が大きく異なる提案手法でも同様の結果が確認できた。ただし、今回の実装では、頭部形状パラメータとして、文献[24]と同じく文献[25]で推定される40次元の顔形状パラメータを利用したが、撮影方向の違いを踏まえると、より適切なパラメータ表現が存在する可能性が考えられる。今回利用した頭部形状パラメータは顔全体の形状を表現できるパラメータであるが、胸部装着カメラから主に観測される部位は鼻と顎であり、それ以外の部分は観測されない。このことから、提案手法では顔全体の形状を表現できるパラメータは不要であり、不要な情報を多く含むことが効率的な個人適応の妨げとなっている可能性が考えられる。そのため、胸部装着カメラによる頭部方向推定に適した形状パラメータを検討するなど、効率的に個人適応が可能な頭部方向推定の実現に向けて検討を進めていく。

4-2 VR 環境における姿勢推定精度の評価

VR 実験環境に関して、まず VR 用トラックを用いた姿勢推定の精度評価を行った。3-4 節で述べた通り、全身に VR 用トラック 9 個を装着し、逆運動学計算用の Unity 用 Asset (Final IK) を利用して全身の各部位の姿勢を推定する。計測の前には、トラックの装着位置に関する Final IK での調整を手作業で行った。

実験では、市販の慣性式モーションキャプチャ (Xsense MNV Awinda) との比較により精度を評価する。図 9 に示すように、VR トラック 9 個と慣性式モーションキャプチャを装着し、両方から得られる姿勢データを比較した。この比較実験では、HMD を装着していないため、HMD の代わりに頭部にも VR 用トラックを装着している。全身の姿勢を推定しているが、ここでは頭部・胸部・腕のみに関する結果を示す。図 10-13 は頭部・胸部の水平方向の角度、腕 (両肘) の角度であり、VR 用トラックでの推定結果、モーションキャプチャシステムでの推定結果と両者の差を示している。横軸は時間である。

結果より、姿勢変化が大きい場合には誤差が大きくなる傾向は見られるものの、基本的には 5 度以内での姿勢推定が可能であることを確認した。今後は、さらなる精度向上を検討するとともに、VR 実験環境における被験者実験を行い、視線方向と身体動作に関するモデル化を行うことで、間接的視線推定を目指していく。



図 9 VR トラックの装着位置

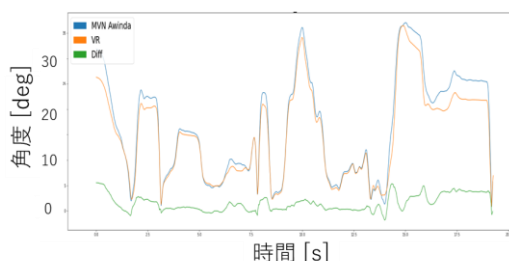


図 11 頭部姿勢推定の比較

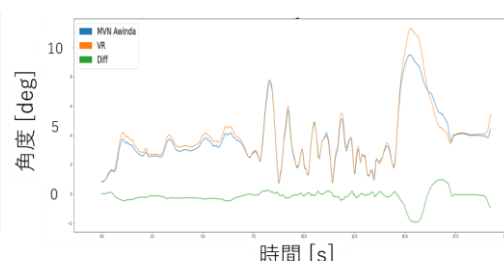


図 10 胸部姿勢推定の比較

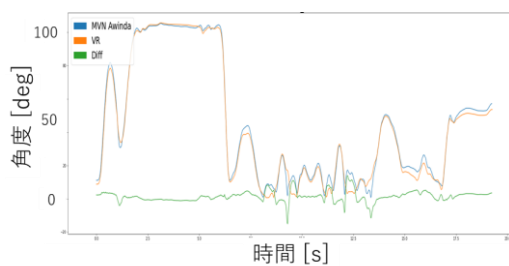


図 12 左肘姿勢推定の比較

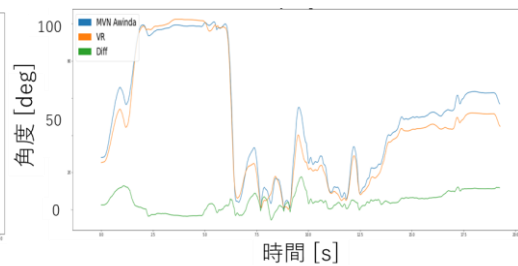


図 13 左肘姿勢推定の比較

5 まとめ

本研究では、視線方向と身体動作の関係性を利用した間接的視線推定を目指し、胸部装着カメラを利用した腕姿勢・頭部姿勢推定と、視線方向と身体動作の関係性のモデル化のための実験環境構築を行った。頭部姿勢推定については、より高精度な姿勢推定を目指し、個人適応が可能な枠組みとし、その有効性を示した。

今後は、新型コロナウイルス感染症の影響で、本研究期間中には実施できなかった被験者実験を行い、視線方向と身体動作の関係性のモデル化を進めるとともに、構築されたモデルに基づく間接的視線推定の評価を行っていく。

【参考文献】

- [1] C. Klein, U. Ettinger (Eds.), “Eye Movement Research”, Springer, 2019
- [2] M. Wedel and R. Pieters, “A review of eye-tracking research in marketing”, Review of Marketing Research, Vol. 4, pp.123-147, 2008
- [3] T. Yonezawa, H. Yamazoe, A. Utsumi, S. Abe, Attractive, “Informative, and Communicative Robot System on Guide Plate as an Attendant with Awareness of User's Gaze”, Paladyn. Journal of Behavioral Robotics, Vol.4, issue.2, pp.113-122, 2013
- [4] L. Fridman, B. Reimer, B. Mehler and W. T. Freeman, “Cognitive load estimation in the wild,” CHI Conference on Human Factors in Computing Systems, Paper No.652, 2018
- [5] T. Fujioka, K. Inohara, Y. Okamoto, Y. Masuya, M. Ishitobi et al., “Gazefinder as a clinical supplementary tool for discriminating between autism spectrum disorder and typical development in male adolescents and adults,” Molecular autism, Vol.7, Article No.19, 2016
- [6] E. D. Guestrin and M. Eizenman, “General theory of remote gaze estimation using the pupil center and corneal reflections,” IEEE Transactions on biomedical engineering, Vol.53, No.6, pp.1124-1133, 2006
- [7] E. Wood, T. Baltrusaitis, L.-P. Morency, P. Robinson, A. Bulling, “Learning an appearance-based gaze estimator from one million synthesised images,” ACM Symposium on Eye Tracking Research & Applications, pp.131-138, 2016
- [8] X. Zhang, Y. Sugano, M. Fritz and A. Bulling, “MPIIGaze: Real-world dataset and deep appearance-based gaze estimation,” IEEE transactions on pattern analysis and machine intelligence, Vol.41, No.1, pp.162-175, 2017
- [9] X. Zhang, Y. Sugano, M. Fritz and A. Bulling, “It's written all over your face: Full-face appearance-based gaze estimation,” IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp.51-60, 2017
- [10] T. Fischer, H. J. Chang, Y. Demiris, “RT-GENE: Real-time eye gaze estimation in natural environments,” European Conference on Computer Vision, pp.334-352, 2018

- [11] T. Shiratori, H. S. Park, L. Sigal, Y. Sheikh, and J. K. Hodgins, “Motion capture from body-mounted cameras,” *ACM Transactions on Graphics*, vol. 30, no. 4, 2011
- [12] W. Xu, A. Chatterjee, M. Zollhoefer, H. Rhodin, P. Fua, H.P. Seidel, C. Theobalt, “Mo2Cap2: Real-time Mobile 3D Motion Capture with a Cap-mounted Fisheye Camera”, *IEEE transactions on visualization and computer graphics*, Vol. 25, No. 5, pp. 2093-2101, 2019.
- [13] D. Tome, P. Peluse, L. Agapito, H. Badino, “xR-EgoPose: Egocentric 3D Human Pose from an HMD Camera”, *International Conference on Computer Vision*, pp.7728–7738, 2019
- [14] H. Yamazoe, A. Utsumi, K. Hosaka, M. Yachida, “A Body-mounted Camera System for Head-pose Estimation and User-view Image Synthesis”, *Image and Vision Computing*, Vol.25, No.12, pp.1848-1855, 2007
- [15] Y. Oozono, H. Yamazoe, J.-H. Lee, “Pointing Direction Estimation for Attention Target Extraction Using Body-mounted Camera”, *Ubiquitous Robots 2020*, 2020.
- [16] D. H. Hwang, K. Aso, Y. Yuan, K. Kitani, H. Koike, “MonoEye: Multimodal Human Motion Capture System Using A Single Ultra-Wide Fisheye Camera,” *ACM Symposium on User Interface Software and Technology*, pp. 98-111, 2020.
- [17] J. Stahl, “Amplitude of human head movements associated with horizontal saccades,” *Experimental Brain Research*, 126, pp.41–54, 1999
- [18] Z. Thumser, B. Oommen, I. Kofman, J. Stahl, “Idiosyncratic variations in eye-head coupling observed in the laboratory also manifest during spontaneous behavior in a natural setting,” *Experimental Brain Research*, 191, pp. 419–434, 2008
- [19] D. Anastasopoulos, N. Ziavra, M. Hollands, A. Bronstein, “Gaze displacement and inter-segmental coordination during large whole body voluntary rotations,” *Experimental Brain Research*, 193, pp. 323–336, 2009
- [20] T. Imai, S.T. Moore, T. Raphan, “Interaction of the body, head and eyes during walking and turning,” *Experimental Brain Research*, 136, pp.1–18, 2001
- [21] H. Yamazoe, I. Mitsugami, T. Okada, T. Echigo, Y. Yagi, “Immersive walking environment for analyzing gaze-gait relations,” *Transactions of the Virtual Reality Society of Japan*, Vol.22, No.3, 2017
- [22] H. Yamazoe, I. Mitsugami, T. Okada, Y. Yagi, “Analysis of head and chest movements that correspond to gaze directions during walking,” *Experimental Brain Research*, Volume 237, Issue 11, pp 3047–3058, 2019
- [23] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *International Conference on Learning Representations*, 2015
- [24] M.Sei et al., “Network Structure for Personalized Face Pose Estimation Using Incrementally Updated Face Shape Parameters”, *ACPR2019*
- [25] X. Zhu, Z. Lei, X. Liu, H. Shi, S.Z. Li, “Face alignment across large poses: A 3d solution,” *IEEE conference on computer vision and pattern recognition*, pp. 146-155, 2016
- [26] A. Quattoni, A. Torralba, “Recognizing indoor scenes,” *IEEE Conference on Computer Vision and Pattern Recognition 2009*, pp. 413-420, 2009