5S-04

アピアランスベースの視線推定における実現可能性を考慮した精度向上

杉山　秀治†　　福田　大翔†　　渡辺　裕†
早稲田大学　基幹理工学研究科†

1. はじめに

視線推定は，人の視線方向を推定する技術である．視線は，人間の非言語コミュニケーションにおいて重要な要素であるため，視線推定はコンピュータビジョンの領域で注目されている．近年では，人間とロボットのインタラクションや自動運転など様々な分野で応用されている．また，視線推定手法の一つにアピアランスベースの手法がある．この手法は，外見画像から直接視線を推定する手法であり，通常の RGB カメラの撮影画像から推定可能であるため，導入のしやすさから現実世界での需要が高い．しかし，外見画像は顔のパーツの個人差や撮影環境の差異を含むため，アピアランスベースによる正確な視線推定は困難である．そのため，高精度の推定を実現するための検討が重ねられている．これまで様々な手法が提案されてきたが，精度改善に焦点が当てられることが多く，現実の応用シーンでの考慮が乏しい．そこで，本論文ではアピアランスベースの視線推定における実現可能性を考慮し，同時に従来手法の精度を向上する手法を提案する．

2. アピアランスベースの視線推定

2.1. 関連研究

アピアランスベースの手法は，Zhang[1]らによる CNN を用いた手法の提案をきっかけに，深層学習の導入により精度の向上を実現してきた．以降深層学習を用いた手法が主流となっている．また，入力画像に関しては，顔画像のみを用いる手法，両目の画像のみを用いる手法，またはこれら両方の画像を組み合わせる手法が存在する．人間の瞳孔の動きは，視線の動きと比べて微小であることから，視線と目の情報は密接な関係にある．加えて，顔に対して目のサイズは小さいため，顔画像のみを用いる場合，目の領域を十分に考慮できていない．したがって，目の情報の粒度を高めるために両目の画像を用いる手法が提案されている．

また，アピアランスベースの手法の一つに L2CS-Net[2]という手法がある．この手法は，顔画像のみを用いる手法ながら，画像認識タスクで広く知られている ResNet をバックボーンとすることで，他の CNN ベースの手法よりも優れた性能を示している． さらに，一般的な視線推定では視線方向をベクトルで推定するのに対して，L2CS-Net では視線方向を三次元座標におけるオイラー角のヨー角とピッチ角に変換し，これらを独立で推定することで精度を向上させている．

2.2. 従来のアピアランスベースの手法の問題点

アピアランスベースの手法では，精度向上のために顔画像に加えて両目の画像を用いるというアプローチをとっている．この手法は精度向上に寄与する一方で，両目の画像を必要とするという問題がある．深層学習による視線推定モデルを構築する際には，顔画像から手動で両目の画像を切り抜くことで，両目の画像を用意することができる．しかし，現実世界でのモデルの応用において，手動での切り抜き作業は導入の際の簡便性を損なう要因になる．このため，何らかの目の検出モデルを前処理として導入する必要があり，視線推定モデルが目の検出モデルの精度に依存するという課題が生じる．したがって，この問題に対して，本研究では両目の画像を必要とせずに視線の推定精度を向上させる手法を目指す．

3. 提案手法

本論文では，アピアランスベースの視線推定において，顔画像のみを用いる精度向上のアプローチを提案する．顔検出モデルは，顔のサイズと目のサイズを比較してもわかる通り，目の検出モデルと比べて，現実世界における高い精度のモデルの実現が可能である．そのため，視線推定においても顔画像を用いてモデルを構築することは，実現可能性を考えても妥当であると考えられる．

Improvement Method for Feasibility in Appearance-Based Gaze Estimation
† Hideharu Sugiyama, Hiroto Fukuta and Hiroshi Watanabe, Waseda University

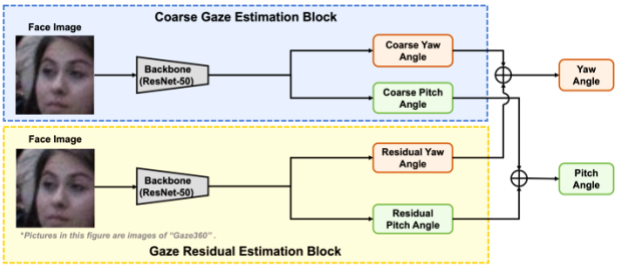

図 1 提案手法の視線推定ネットワーク

我々の提案手法は，顔の全体に注目するブロックと，視線と密接な目の領域に注目するブロックの二つの要素で構成されている．提案手法の視線推定ネットワークを図 1 に示す．

まず，青色領域の Coarse Gaze Estimation Block では，顔画像全体から顔の方向を含む，大まかな視線方向の角度（Coarse Angle）を推定する．次に，黄色領域の Gaze Residual Estimation Block では，顔画像から Coarse Angle を正解の視線方向に補正する角度（Residual Angle）を推定する．このブロックにより，視線の変化による瞳孔や目の微細な動きをとらえることが可能となる．

提案手法の構成は，人間の視線メカニズムに基づいている．人間は対象物に視線を向ける際，頭を動かして対象物に顔を向け，目を動かすことで焦点を合わせることが一般的である．そのため，この仕組みに注目し，視線方向が顔の向きを中心としていることと，瞳孔の微細な動きによる視線の変化を捉えることが重要と考え，二つのブロックで人間の視線メカニズムをモデル化する．この提案手法により，顔画像のみを使用しながらも瞳孔の動きをとらえ，視線推定の精度向上を目指す．

4. 実験

評価実験には，Gaze360[3]，RT-GENE[4]の二つのデータセットを用いる．特に Gaze360 は他のデータセットよりも視線方向の分布が広く，よりオープンな環境を想定している．評価指標には，平均角度誤差，Mean Angular Error（MAE）を用い，式(1)に示す．$\hat{\boldsymbol{g}}$は推定した視線方向，$\boldsymbol{g}$は正解の視線方向であり，値が小さいほど良い性能を示す．

$$MAE = \frac{1}{N}\sum_{i=1}^{N} \arccos\left(\frac{\hat{\boldsymbol{g}}\cdot\boldsymbol{g}}{\|\hat{\boldsymbol{g}}\|\|\boldsymbol{g}\|}\right). \quad (1)$$

表 1 は，提案手法と従来手法の比較結果である．また，Gaze360 が広い視線分布を持つことから，0~180°，0~90°，0~60°といった，正解の視線方向のレンジを三段階用意し，これら評価セットに対しても MAE を求めた．これら結果を表 2 に示す．表 1，2 ともに MAE の単位は°である．

表 1 提案手法と従来手法の実験結果

手法	データセット	
	Gaze360	RT-GENE
L2CS-Net	10.42	6.59
Ours	**10.19**	**6.55**

表 2 Gaze360 における提案手法と従来手法の比較

手法	Gaze360		
	0~180°	0~90°	0~60°
L2CS-Net	10.42	9.29	9.07
Ours	**10.19** (-0.23)	**9.04** (-0.25)	**8.70** (-0.37)

表 1 より，提案手法が従来手法に対して性能向上を示すことがわかる．これより，提案手法がアピアランスベースの視線推定において有効であることが示唆される．ただし，被験者の少ない RT-GENE では改善幅が小さいことから，被験者数が多い場合に提案手法がより有効である可能性が伺える．また，表 2 より提案手法が被験者の視線方向が狭くなるにつれて，性能向上を示すことがわかる．これより，被験者の顔がカメラの正面方向となるケースに，より有効な手法であることがわかる．

5. まとめ

本論文では，アピアランスベースの視線推定において，実現可能性を満たしながら精度を向上させる手法として，顔画像のみを用いた二つのブロック構成による手法を提案した．評価実験により，提案手法が従来手法に対して性能向上を示すことを確認した．我々の提案手法は，顔画像のみを使用するアピアランスベースの視線推定手法に適用可能であり，広い活用が期待される．

参考文献

[1] Zhang, X., Sugano, Y., Fritz, M. and Bulling, A., “Appearance-Based Gaze Estimation in the Wild,” in The IEEE Conference on Computer Vision and Pattern Recognition, 4511–4520 (2015).
[2] Abdelrahman, A. A., Hempel, T., Khalifa, A. and Al-Hamadi, A., “L2CS-Net: Fine-Grained Gaze Estimation in Unconstrained Environments,” arXiv preprint arXiv:2203.03339, (2022).
[3] Kellnhofer, P., Recasens, A., Stent, S., Matusik, W. and Torralba, A., “Gaze360: Physically unconstrained gaze estimation in the wild,” in The IEEE International Conference on Computer Vision, 6912–6921(2019).
[4] Fischer, T., Chang, H. J. and Demiris, Y., “RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments,” in the European Conference on Computer Vision, 339–357(2018).