令和 7 年度 卒業論文

視線推定と非接触脈拍推定を統合した集中状態推定システムの提案

指導教員　岡部 誠 准教授

提出者　50216014 小笠原 侑紀

提出日　2026 年　2 月　12 日

目次

第1章　はじめに

現代の高度情報化社会において，学習や業務における生産性を維持・向上させるためには，高い集中状態を持続することが極めて重要である.集中状態は作業成績や学習効率に直結する一方で，作業者が自らの集中度を客観的かつリアルタイムに把握し，適切にセルフマネジメントを行うことは容易ではない.一般に集中力は主観に依存する側面が大きく，作業者自身が「集中している」と感じていても，実際には注意が散漫となりパフォーマンスが低下している場合も少なくない.したがって，客観的なデータに基づき集中状態を可視化し，作業者が自身の状態変化を振り返ることのできる外部支援システムの構築が求められている.さらに，集中状態の可視化は本人の支援に留まらず，第三者による活用にも意義がある.例えばプレゼンテーションや発表の場面では，話し手が聴衆の理解度や関心を推し量りながら進行を調整することが望ましいが，聴衆が「集中して聞いているか」を客観的に把握することは難しい.もし聴衆の集中状態を定量的に推定できれば，話し手や指導者は説明速度や資料提示のタイミングを調整でき，発表の質や学習効果の向上に繋げられる.よって，集中推定はセルフマネジメント支援のみならず, 教育や対人コミュニケーションにおける第三者支援としても有用である.

集中状態を推定する手法としては，脳波（EEG）や装着型センサーによる脈拍計測など，高精度な生体情報計測に基づく方法が数多く提案されてきた（Weixuan Chen と Daniel McDuff[14] , Zitong Yu ら[16], [17]）.しかし，これらの手法はデバイス装着に伴う身体的拘束感や違和感という問題を抱えている.装着行為そのものが作業者の集中を妨げる要因となり得る点は本質的な欠点であり，本来計測したい自然な集中状態を歪める可能性がある.ゆえに，日常環境への導入や長時間利用を想定する場合，非接触かつ低コストに状態推定を行える枠組みが必要である.一方で，カメラ映像を用いる手法では，一般的な安価な Web カメラにおいて解像度やフレームレートが不足し, 眼球運動のような微細な変化を安定して捉えることが難しい.そこで本研究では，スマートフォンを高画質な Web カメラとして利用可能な iVCam を活用し，高精度な映像解析が可能な計測環境を構築する.

本研究では，コンピュータビジョン技術を用いて，眼球運動の時系列変化および映像から推定される脈拍情報を統合し，集中状態を推定するシステムを提案する.本システムは特別な専門機材を必要とせず，スマートフォンカメラと Python 環境により実装可能であり，OpenCV および MediaPipe を用いて顔情報を抽出し, L2CS-Net により視線方向を Yaw 角・Pitch 角として推定する.さらに，連続フレーム間の角度変化から視線速度を算出し，視線の安定性を定量化する.加えて，顔領域の微細な色情報変化を解析して非接触型の脈拍推定を行い, 視線情報と組み合わせて多角的な状態推定を目指す.以下の図 1(a),図 1(b)に本システムの出力結果を示す. 図 1(b)の時系列ログを見てみると、例えば、開始から 2.2 秒間は視線が正面で安定し脈拍も平常であるため「FOCUSED（集中)」と表示されるが、その後の 4.2

秒間で視線が大きく右側に逸れた（よそ見）場合には、システムが即座に「DISTRACTED（注意散漫)」へとラベルを切り替え、バウンディングボックスを赤色に変化させて警告する。提案システムの特徴は三つである.第一に，高画質映像に基づく非接触・非侵襲な計測である.第二に，眼球運動速度と脈拍を統合した状態推定である.第三に，推定結果の時系列可視化によるフィードバック支援である.使用者は自身の集中の波を客観的に振り返ることができ，休憩の最適化や作業環境改善などのセルフマネジメントに活用できる.さらに第三者がこの情報を参照することで，発表や講義において聴衆の集中が高い区間と低い区間を推定し，説明方法や進行の改善に役立てることができる.以上より，本研究で提案するシステムは，個人のパフォーマンス向上のみならず，対人場面におけるコミュニケーションの質を高める支援技術としても有用であると期待される.

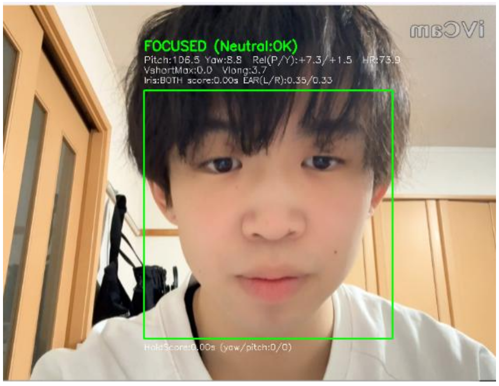


図 1(a) 本システム実行中の画面

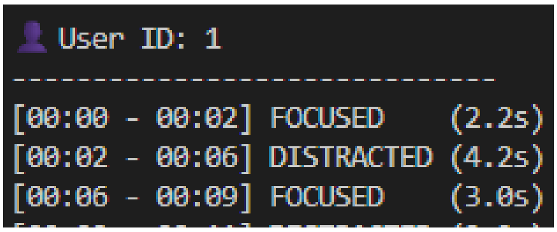


図 1(b) 時系列ログ

第2章　関連研究

2.1 MediaPipe による顔面情報の抽出

非接触での状態推定において，顔面のランドマーク（特徴点）を高精度かつリアルタイムに取得することは，視線推定および脈拍推定の双方の起点となる極めて重要な工程である.従来, Dlib 等の古典的手法や初期の CNN ベース手法でも顔特徴点の推定は可能であったが，実環境では照明変動・頭部姿勢変化・部分遮蔽などの外乱が生じるうえ，視線推定や rPPG のような重い処理と併走させる場合，処理遅延が蓄積しやすい.したがって，本研究のように「視線速度の変化」や「脈拍の変動」を即座に捉える用途では，軽量で安定した顔情報抽出基盤の採用が不可欠である.本研究では，Google が提供するオープンソースのパイプラインフレームワークである MediaPipe を採用する.MediaPipe は，認識・前処理・後処理をグラフ(パイプライン)として構成し, 各モジュールを効率よく連結できる設計を持つ(Camillo Lugaresi ら[1]）.この枠組みにより，顔検出，ランドマーク推定，ROI 抽出などを段階的に

処理し，フレームごとの負荷を抑えやすい.また，MediaPipe はモバイル動作を想定した軽量設計であり，リアルタイム処理に適している点が実用上の利点となる（Camillo Lugaresi ら[1]）.さらに，顔検出器として BlazeFace が提案されており，モバイル GPU を想定した高速な顔検出を目指す方向性が示されている（Valentin Bazarevsky ら[2]）.本研究では，MediaPipe の Face Mesh を用いて顔ランドマークを取得し，目周辺や皮膚領域の ROI を安定に追跡する基盤として用いる.Face Mesh は高密度なランドマークにより顔形状を細かく表現でき，目周辺形状（閉眼判定に関わる特徴）や皮膚領域の追跡（脈拍推定 ROI）を一貫した座標系で扱える（Camillo Lugaresi ら[1]）.また，顔解析ツールキットとして OpenFace 2.0 が提案され，顔行動解析の実験基盤として有用であることも示されている（Tadas Baltrušaitis ら[3]）.本研究は MediaPipe を採用するが，このような実験基盤の知見を踏まえ，顔情報取得の安定性が後段の推定結果の信頼性を左右する点を重視する.以上より，本研究では MediaPipe を，作業者の身体を拘束せずに顔情報をリアルタイム取得する基盤として利用する.図 2 に MediaPipe による顔ランドマーク出力例を示す.

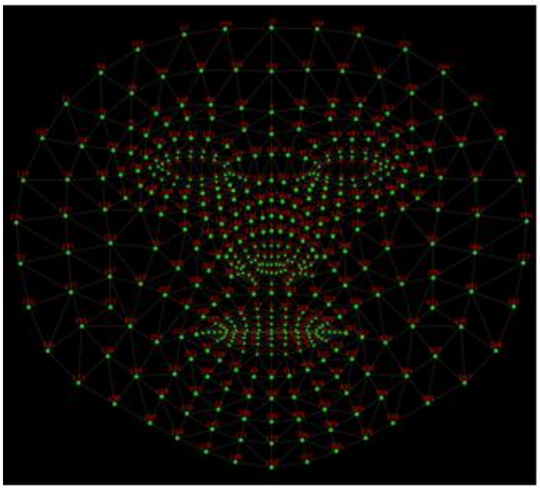


図 2 MediaPipe による顔面のランドマーク抽出結果.

2.2 L2CS-Net による視線推定

本研究では，集中状態と関連が深いと考えられる「視線の動き」を定量化するため，視線方向を Yaw/Pitch 角として推定し，さらに角度系列の時間変化から視線速度を算出する.従来の視線推定は，専用のアイトラッカーなどを用いた高精度計測が可能である一方，作業環境で常時装着することは負担となり得る.そこで近年は，単一カメラから顔画像を入力し，視線方向を推定する外観（appearance）ベース手法が広く研究されてきた.代表例として，スマートフォン等の一般カメラでの視線推定を目指した iTracker が提案され，汎用デバイス上での視線推定の可能性が示されている（Kyle Krafka ら[4]）.また，日常環境下での視線推定のために MPIIGaze が提案され，実環境のばらつきを含むデータセット整備と評価の

重要性が示された（Xucong Zhang　ら[5]）.さらに，より非拘束環境（物理的制約の少ない条件）を扱う Gaze360 が提案され，頭部姿勢や撮影条件が変動する状況での視線推定が議論されている（Petr Kellnhofer　ら[6]）.極端な頭部姿勢や視線変化を含む大規模データセットとして ETH-XGaze も報告され，汎化性能評価の基盤として位置づけられる（Xucong Zhang　ら[7]）.また，実時間性を意識した視線推定として RT-GENE が提案され，自然環境下でのリアルタイム推定の難しさと対処が議論されている（Tobias Fischer　ら[8]）.本研究で採用する L2CS-Net は，顔画像から Yaw/Pitch を推定する深層学習モデルであり，角度推定を分類と回帰の枠組みで扱うことで精度向上を狙う（Ahmed Abdelrahman　ら[9]）.本研究において L2CS-Net を用いる利点は次の 3 点である.第一に，Yaw/Pitch がフレームごとに連続量として得られるため，後段で視線速度（角度変化量）へ自然に接続できる（Ahmed Abdelrahman　ら[9]）.第二に，頭部姿勢・照明などの外乱を含む「非拘束環境」を想定した文脈に位置づけられ，実環境での利用を前提とした議論と整合する（Petr Kellnhofer　ら[6]，Tobias Fischer　ら[8]）.第三に，視線推定の出力が時系列として得られることで，単一フレームの角度そのものだけでなく，「短時間の急激な揺れ」「長時間の落ち着きのなさ」といった時間的特徴を定義できる.一方で，Yaw/Pitch 角はカメラ設置位置や顔の向き，姿勢の個人差により一定のオフセットを含みやすい.この問題に対して，本研究では「neutral（正面基準)」を導入し，相対角として扱うことで判定の一貫性を確保する.以降の状態推定では角度そのものを主指標とせず，角度系列の変化量に基づく視線速度を主要指標として扱う.図 3 に L2CS-Net の構造概略を示す.

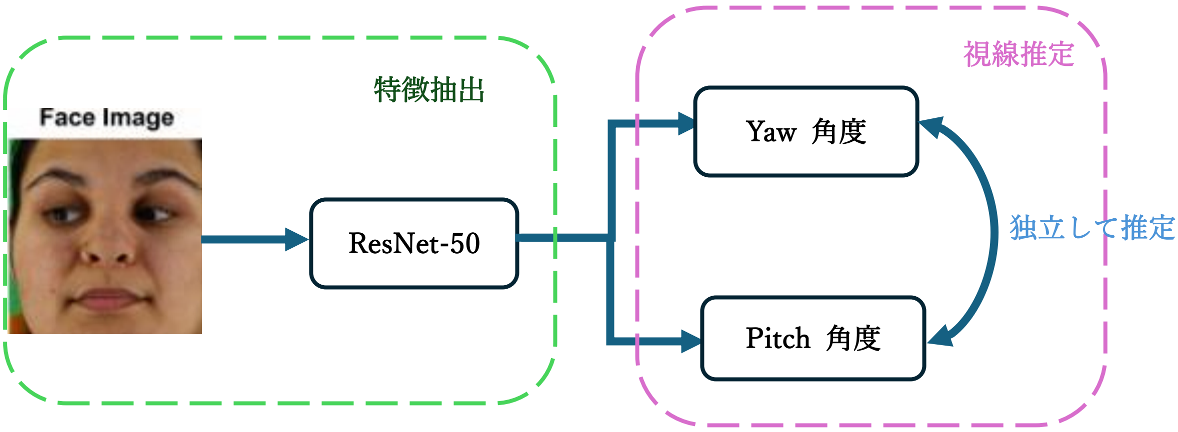


図 3　L2CS-Net の視線推定ネットワークの構造

2.3 Eulerian Video Magnification による脈拍解析

本研究のもう一つの柱は，顔映像から脈拍（BPM）を非接触で推定することである.接触型センサ（心電計やパルスオキシメータ）を用いれば高精度な計測が可能だが，装着の煩わしさや身体拘束による不快感が，作業中の状態に影響を与える可能性がある.したがって，本研究では映像のみを用いた非接触計測（remote PPG: rPPG）を採用する.映像から脈拍を

推定する際，皮膚表面の血流に伴う微小な反射変化は非常に弱く，照明変動や微小な頭部運動の影響を受けやすい.この「微小信号を扱う」課題に対して，Eulerian Video Magnification（EVM）（図 4）は重要な関連技術である.Hao-Yu Wu ら[10]は，時間方向のフィルタリングと増幅を用いて，映像中の微細変化を可視化する EVM を提案した.Neal Wadhwa ら[11]は EVM の考え方と応用を整理し，微小信号の抽出・解析の観点をまとめている.EVM の典型的な構成は，(i）空間分解（ピラミッド分解），(ii）時間方向の帯域通過フィルタリング，(iii）成分の増幅と再合成，であり，心拍周波数帯域に相当する周期成分の抽出と親和性が高い（Hao-Yu Wu ら[10]，Neal Wadhwa ら[11]）.さらに，EVM 派生の関連として，位相情報に基づく動画処理（Phase-based motion processing）が報告され，微小な動きの抽出を扱う枠組みが提示されている（Neal Wadhwa ら[12]）.また，大きな運動が存在する状況での動画増幅についても検討があり，実環境での頑健性が議論されている（Mahmoud Elgharib ら[13]）.ただし，本研究の目的は「色変化の可視化」そのものではなく，脈拍（BPM）として安定に数値化し，状態推定へ統合することである.そのため，近年の深層学習ベース rPPG の知見も参照する.Weixuan Chen と Daniel McDuff[14]は DeepPhys を提案し，畳み込み注意機構により生理信号に寄与する領域を強調して推定する枠組みを示した.Xin Liu ら[15]はオンデバイス計測を意識した時系列シフトと注意を組み合わせた手法（MTTS-CAN）を提案し，計算効率と精度の両立を議論している.Zitong Yu ら[16]は時空間ネットワークによる rPPG 推定を扱い，時系列表現の重要性を示した.Transformer 系として PhysFormer も提案され，時系列差分に着目した表現学習が報告されている（Zitong Yu ら[17]）.さらに実用面では，処理を簡素化しつつ高速化を狙う EfficientPhys が提案され，実時間性の観点から有用である（Xin Liu ら[18]）.

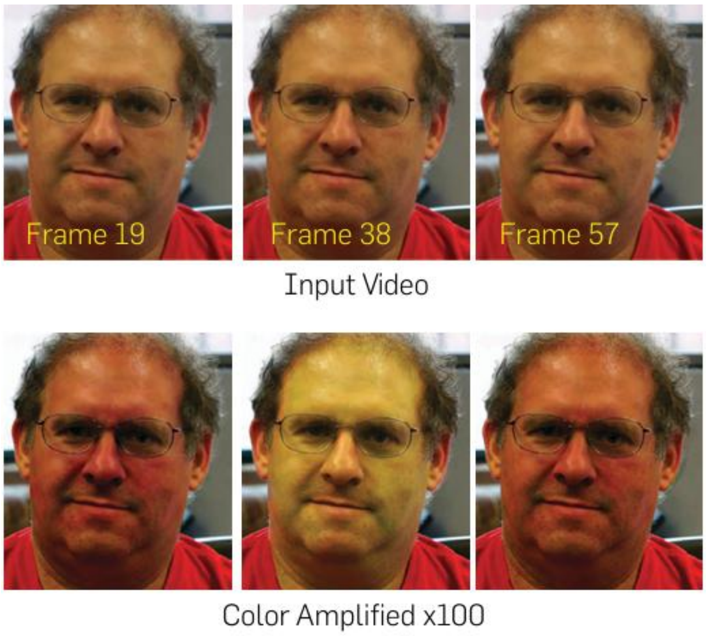


図 4　EVM による入力映像から特定の周波数帯域（心拍成分）を抽出し、増幅する過程

2.4 視線と生体信号を用いた状態推定

集中・注意散漫・眠気・緊張などの状態は，単一指標だけで確定することが難しい.例えば，視線だけでは「必要な探索行動」と「落ち着きのない視線（きょろきょろ)」の区別が曖昧になり得る.一方，脈拍だけでは姿勢変化，呼吸，周囲環境など作業外要因による変動を除外しにくい.そこで近年は，複数モダリティを統合して推定の頑健性を高める設計思想が広く採用されている.特に運転者モニタリング領域では，実環境の外乱下で状態を安定に推定するため，視覚情報と生体情報を併用する流れがある.例えば，カメラベースの生体計測を運転者モニタリングへ応用する観点から SparsePPG が報告され，照明変動や運動が存在する状況での課題が整理されている（E. M. Nowara ら[19]）.また，リアルタイム rPPG そのものの軽量化・安定化に向けた検討も進んでおり，実用的な状態推定へ接続する際には「推定の安定性」と「処理遅延」の両面が重要になる（Xin Liu ら[15]，[18]）.このような文脈を踏まえると，視線・脈拍の統合では「どちらか一方の指標が一時的に乱れた」だけで状態を切り替えるのではなく，複数条件が整合したときに状態を確定することが実用上有利である.本研究は，上記の潮流を踏まえつつ，計測装置を増やさずに導入可能な最小構成として，視線（相対 Yaw/Pitch＋視線速度）と脈拍（BPM）を統合する.統合の基本方針は，(i) 視線の信頼性（両目開眼・虹彩取得の安定性）を先に確認し，(ii) 視線方向の継続逸脱（よそ見）と視線速度（きょろきょろ）を時間窓で評価し，(iii) 脈拍は基準からの変化量として補助的に用い，複数条件が同時に成立した場合にのみ状態を確定する，というものである.これにより，視線のみ・脈拍のみの単独変化に起因する誤判定を抑え，実環境における外乱や個人差を含む条件でも安定した状態推定を目指す.

第 3 章 提案手法

3.1 システムの概要

図 5 は，本システムの全体構成と情報の流れを示す概要図である.左側に示すように，入力は iVCam を介して PC へ取り込まれるスマートフォンのカメラ映像であり，PC 側ではこの映像を連続フレームとして取得して以降の処理に用いる.中央のブロックは MediaPipe Face Mesh による顔情報抽出であり，フレームごとに顔のランドマークを検出することで，顔位置の追跡と目周辺を含む解析の基盤を形成する.ここで得られた顔情報を起点として，処理は上段の視線推定と下段の脈拍推定の二経路に分岐する.上段では，顔画像（または目領域）から視線方向を推定し，Yaw および Pitch として出力することで，視線の向きや変化を表現する.下段では，顔領域内の一部を ROI（関心領域）として設定し，その領域の画素値変動を解析することで非接触型の脈拍を推定する. 右端には出力結果の例を示しており，推定された状態ラベルに加えて，視線・脈拍などの推定値が映像上に重畳表示される.さら

に，推定結果は時系列ログとしても出力されるため，状態変化を後から区間単位で確認できる．

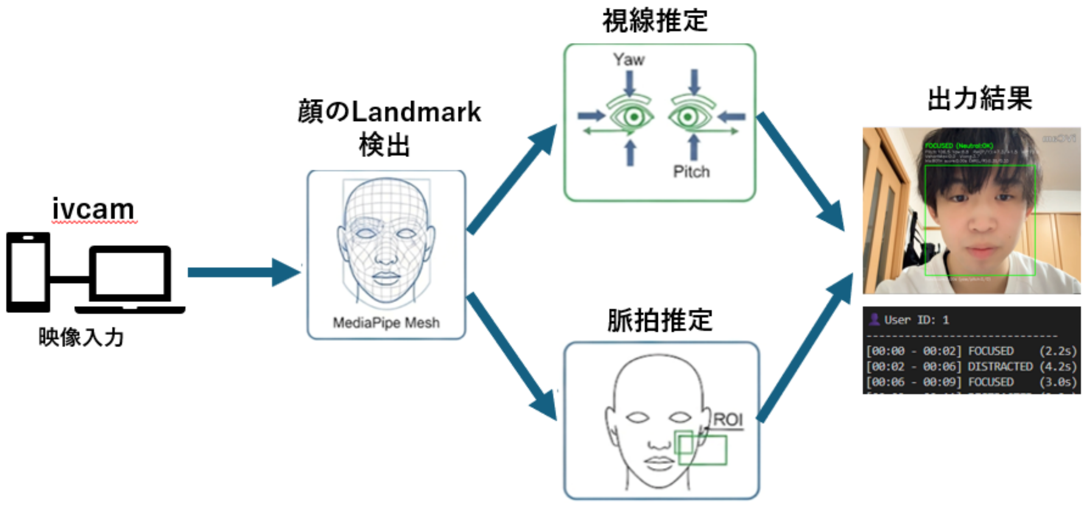


図 5 本システムの概要図

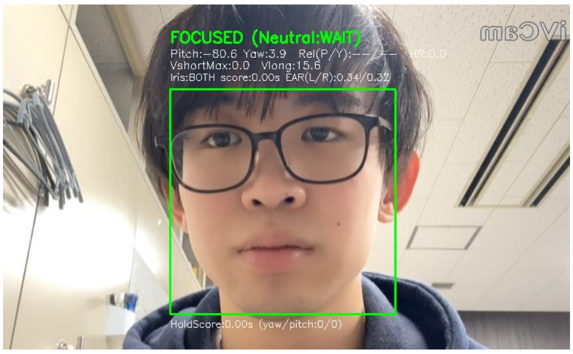
3.2 基準値の設定

本システムでは，視線推定および脈拍推定によって得られる時系列データから特徴量を算出し，それらに基準値を設けることで使用者状態を評価する．視線については，注視方向そのもの（Yaw 角・Pitch 角）に加え，連続フレーム間の変化量に基づく視線速度を主要指標として用いる．これは，集中状態の低下が「視線の大きな逸れ」だけでなく，「細かな視線の動きの増加」としても現れるためである.脈拍については，顔映像から推定される BPM を用い，過度な変動や基準からの乖離を状態推定の補助情報として利用する．

3.2.1 視線推定における基準値

視線推定では，L2CS-Net が出力する Yaw 角（左右）および Pitch 角（上下）を用いて，使用者がどの方向を注視しているかを表現する.しかし，カメラの設置位置や顔の向き，個人の姿勢によって，推定される角度には一定のオフセットが生じる.このため，本システムでは Yaw/Pitch の絶対値をそのまま判定に用いず，「正面の基準（neutral)」を導入し，そこからの相対角として扱う（図 6).図 6 左は neutral がまだ確定していない状態を示しており，表示は「Neutral: WAIT」となる.一方，図 6 右では neutral が確定した状態を示しており，「Neutral: OK」と表示され，この基準からの相対角が算出される.以降の状態判定は，相対角に基づいて行うことで，個人差やカメラ条件が異なる環境でも判定の一貫性を保てる．

さらに,neutral は視線が安定している区間の Yaw/Pitch を蓄積して推定する.



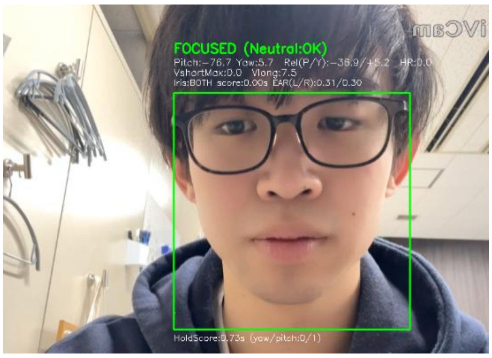


図 6 左：Neutral: WAIT　　　図 6 右：Neutral: OK

Yaw 角・Pitch 角の基準値は,「正面から大きく逸れている状態」を判定するために用いる.例えば,|ΔYaw|が 25° 以上の状態が 1.0 s 以上継続する場合は「向きっぱなし」とみなし,注意散漫の兆候として扱う.Pitch 角についても同様に,|ΔPitch|が 20° 以上の状態が 1.0 s 以上継続する場合は,注視対象から外れている可能性が高いと判断する.ただし,本研究の状態推定では角度そのものを主指標とはせず,次に述べる視線速度（V_{gaze}）をより主要な指標として扱う.視線速度は,連続フレーム間の Yaw/Pitch の変化量から算出する.フレーム間隔を Δt,Yaw と Pitch の変化量をそれぞれ ΔYaw,ΔPitch とすると,視線速度 V_{gaze} は Yaw–Pitch 平面上の移動量として次式で定義できる(図 7).

$$V_{gaze} = \sqrt{\left(\frac{\Delta Yaw}{\Delta t}\right)^2 + \left(\frac{\Delta Pitch}{\Delta t}\right)^2}$$

この指標により,視線がどれだけ速く動いているかをフレームごとに数値化できる.ただし,推定誤差による角度の急跳びが生じると V_{gaze} が過大化し,状態判定が不安定になる.そこで本研究では,|ΔYaw|>25° または |ΔPitch|>25° を満たすフレームを無効サンプルとして除外し,速度指標の安定性を確保する.

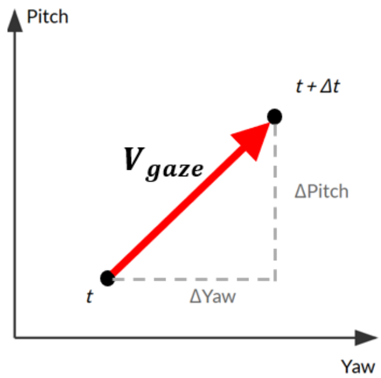


図 7 視線速度の表し方

単一の時間幅のみでV_{gaze}を評価すると,「瞬間的なキョロキョロ」と「持続的なそわそわ」を区別しにくい.そこで本システムでは，短い時間幅と長い時間幅の両方でV_{gaze}を評価する.短い時間幅として直近 0.4 s の速度最大値 V_{short} を用い，これが 40 deg/s を超えた場合を「瞬間的な揺れ」とみなす.一方，長い時間幅として直近 2.0 s の速度平均 V_{long}を用い，これが 15 deg/s を超える状態が 2.0 s 以上継続する場合を「持続的な揺れ」とみなす.これにより，短時間の視線移動を過大に「注意散漫」と誤判定することを抑えつつ，持続的に落ち着きのない視線を検出しやすくなる.なお，これらの閾値（40 deg/s，15 deg/s）は予備計測ログに基づく初期設定であり，環境や被験者に応じて取得ログから微調整可能な設計とする(図 8).

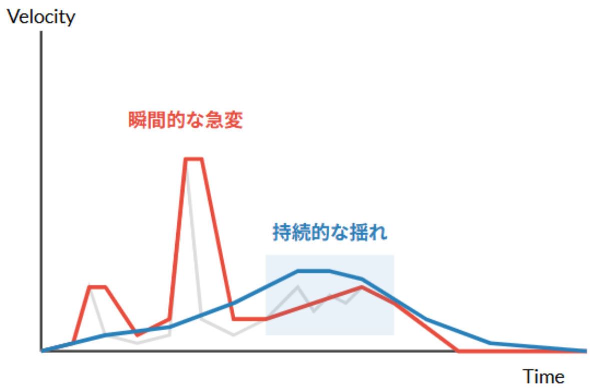


図 8 単一時間幅での評価

3.2.2 脈拍推定における基準値

脈拍推定では, 顔領域のうち血流変化が比較的安定して観測できる領域を解析対象（ROI）として設定し，その画素値変動から BPM（心拍数）を算出する.ただし，映像由来の脈拍推定値は照明変動や微小な頭部運動の影響を受けやすく，生データにはノイズが含まれる.そこで本研究では，図 9 に示すように生データ（灰色）に対して移動平均などの平滑化処理を適用し，急激な跳ねを抑えた平滑化 BPM（赤）を評価に用いる.これにより，一時的な外乱による誤ったピークを抑制し，状態変化の傾向を安定して捉えやすくする.基準値の設定は，被験者ごとの安静時心拍に大きな個人差が存在する点を踏まえ，絶対値ではなく「基準からの変化率」を中心に扱う方針とする.具体的には，図 9 の①に示す初期区間を基準推定区間として設定し，この区間の BPM から各被験者の基準 BPM（破線）を推定する.本システムでは，基準から 20% 以上の低下が観測される場合を「覚醒度の低下（眠気）の兆候」として扱う（図 9③）.一方で，基準から 7% 以上の上昇が観測される場合を「ストレス・緊張状態の兆候」として扱う（図 9②）.このように，個人ごとの基準値に対する割合で判定することで，一律の BPM 閾値を用いた場合に生じやすい誤判定を低減できる.ただし,脈拍は作

業内容や姿勢，呼吸など多様な要因でも変動し得るため，本研究では脈拍のみで状態を確定せず，視線速度などの視線指標と統合して最終的な状態推定に用いる．

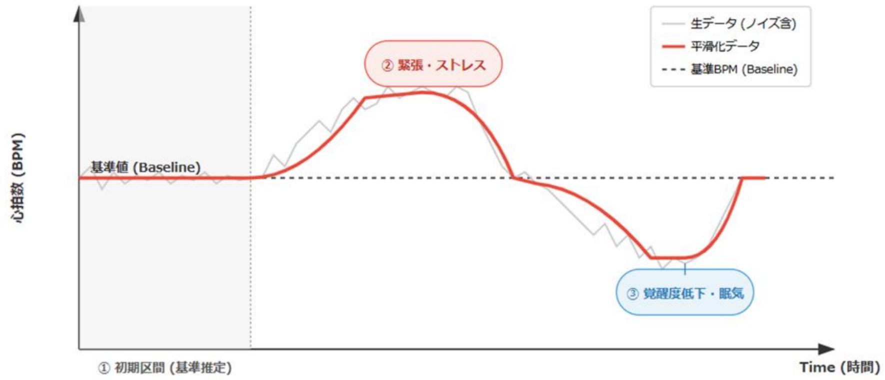


図 9 脈拍に基づく評価

3.2.3 視線推定と脈拍の統合評価

本システムでは，視線（相対角 Yaw/Pitch，視線速度 V_{gaze}）と脈拍（BPM）を用いて使用者状態を推定する（図 10）.判定は図のフローに従い，上から順に条件を確認し，いずれか一つでも NG であれば「注意散漫」，すべて OK であれば「集中」とする．

まず，両目が開いているかどうかを判定する.瞬きなど一時的な閉眼は除外しつつ，閉眼が継続して検出された場合は以降の視線推定が不安定になるため NG とし，「注意散漫」と判定する．

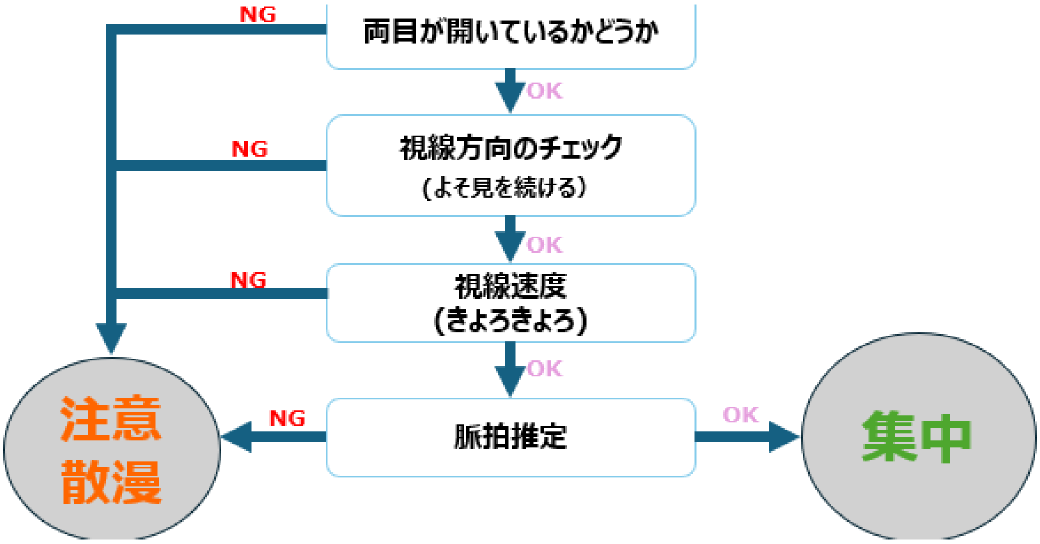
次に，視線方向のチェック（よそ見を続ける）を行う.ここでは neutral を基準とした相対角（Yaw/Pitch）を用い，個人差やカメラ設置条件に起因するオフセットの影響を抑えた上で，「正面から大きく外れた状態」が継続していないかを確認する.所定の条件を満たすよそ見が継続した場合は NG とし，「注意散漫」と判定する．

続いて,視線速度（きょろきょろ）を評価する.視線速度 V_{gaze} は連続フレーム間の Yaw/Pitch 変化量から算出し，短時間窓（0.4 s）の最大値と長時間窓（2.0 s）の平均値の両方で判定する.具体的には，短時間窓の最大速度が 40 deg/s を超える場合を「瞬間的な揺れ」，長時間窓の平均速度が 15 deg/s を超える状態が継続する場合を「持続的な揺れ」とみなし,いずれかが成立した場合は NG として「注意散漫」と判定する．

最後に，脈拍推定を用いて状態を補強する.脈拍は平滑化 BPM を用い，被験者ごとの基準 BPM に対する変化率で評価する.本システムでは, 基準から 7% 以上の上昇をストレス・緊張の兆候，20% 以上の低下を眠気（覚醒度低下）の兆候として扱う.これらの変動が観測

される場合は，集中状態とはみなしにくいため NG として「注意散漫」と判定する.一方，脈拍が基準近傍で大きな変動が見られない場合は OK とし，ここまでの全チェックが OK であった場合に限り「集中」と判定する.

また，リアルタイム処理ではラベルが短時間で頻繁に切り替わると可読性が低下するため，本システムでは状態遷移に保持時間を設け，判定が成立した場合は一定時間その状態を維持することで，表示の揺れを抑制する.



10 統合判定のフロー

3.3 実装

3.3.1 実行環境

本システムは，Windows 11 を搭載した Intel Core i7（第 11 世代）ノートパソコン上で実行した.メモリは 16GB であり，GPU は非搭載または内蔵 GPU 環境を想定し，主要な処理は CPU 上で動作する構成である.実装言語は Python 3.10 である.映像入力には，画質と入力の安定性を確保するため Web カメラではなく iVCam を用い，スマートフォンのカメラ映像を PC へ入力する.PC 側では OpenCV を用いて映像を逐次取得し，フレーム列として後続の顔検出・視線推定・脈拍推定へ供給する.主要ライブラリとして，顔検出およびランドマーク推定に MediaPipe，視線推定に L2CS-Net（PyTorch），信号処理に NumPy 等を用いる.なお，処理は入力フレームレートに可能な限り追従するリアルタイム動作を目標とし，推定結果は後述の評価に利用できるようログとして保存可能な形で出力する.

3.3.2 オンライン処理の制御

リアルタイム性を確保するため，本システムでは「毎フレーム実行する処理」と「一定間隔で間引いて実行する処理」を分離する設計とした.まず，各フレームに対して MediaPipe を適用し，顔領域を検出して切り出し画像を生成する.顔バウンディングボックスはフレームごとの検出揺れを含むため，平滑化を施して切り出し位置の安定化を図る.顔が一時的に検出できない場合は，推定更新を抑制し表示のみ継続することで，誤検出に起因する状態ラベルの乱れを低減する.視線推定では，切り出した顔画像を L2CS-Net へ入力し，Yaw 角（左右）および Pitch 角（上下）を推定するが，推論負荷を抑える目的で毎フレームではなく一定間隔で実行する.推定角度系列はノイズを含むため，平滑化を適用して時系列の安定性を高める.脈拍推定では，顔領域内の一部を解析対象領域（ROI）として設定し，ROI の画素値変動から時間系列信号を構成する.信号は一定時間窓ごとに周波数解析（FFT）を行い，心拍に相当する成分を BPM として推定する.これらの処理における窓長や更新周期，帯域設定などのパラメータは，実装上の処理手順として本節では概要に留め，基準値や設定根拠は別節で述べる.

3.3.3 出力と記録

本システムは，推定した Yaw 角・Pitch 角，視線指標（例：視線速度），脈拍（BPM），および状態判定結果をリアルタイムに画面へ重畳表示し，利用者がその場で状態を確認できるようにする.また，区間ごとの状態ラベルや推定指標の時系列をログとして保存し，後段の評価実験やユーザスタディにおいて，課題成績や主観評価との関係を分析可能なデータとして蓄積する.これにより，単に推定結果を提示するだけでなく，状態推定の妥当性検証と改善に必要な情報を体系的に取得できるようにした.

第 4 章 結果と考察

4.1 出力結果

本システムは，入力映像に対して顔検出を行い，視線推定および脈拍推定に基づく状態判定をリアルタイムに出力する.図 11 に示すように，画面左上には現在の状態ラベル（FOCUSED / DISTRACTED / DROWSY）が表示され，判定に用いた主要特徴量として，Yaw 角・Pitch 角，正面基準（neutral）からの相対角，視線速度（V_{short},V_{long}），虹彩取得状況（Iris），閉眼判定に関連する EAR，および脈拍（HR）が同時にオーバレイ表示される.さらに，顔領域にはバウンディングボックスが描画され，FOCUSED 時は緑，DISTRACTED

時は赤，DROWSY 時は青に切り替わることで，状態の変化を直感的に把握できる出力形式となっている.状態判定は，誤判定を抑えるために図 10 のフローに従う直列判定として実装する.すなわち，①両目が開いているか（Iris/EAR），②視線方向が正面から大きく逸脱していないか（Rel），③視線が落ち着きなく移動していないか（V_{short},V_{long}），④脈拍が平常時基準から大きく変化していないかを上から順に確認し，いずれかが NG であれば注意散漫側（DISTRACTED / DROWSY）とする.視線速度については，短時間窓（0.4 s）の最大値 V_{short} が 40 deg/s を超える場合を瞬間的な揺れ，長時間窓（2.0 s）の平均値V_{long} が 15 deg/s を超える状態が継続する場合を持続的な揺れとして扱い，これらが成立した場合は「きょろきょろ」とみなして注意散漫の判定根拠とする.脈拍については被験者ごとに基準 BPM（平常時）を推定し，基準からの変化率で評価する.本システムでは，平常時から 7% 上昇した場合をストレス・緊張の兆候，20%低下した場合を眠気（覚醒度低下）の兆候として扱い，最終判定に反映する.図 11(a)は FOCUSED（Neutral: OK） と判定された例である.正面を基準とした相対角が小さく，視線速度（V_{short},V_{long}）も低値で推移しており，Iris が BOTH となって両眼情報が安定して取得できている.これらの条件が連続して成立した場合，本システムは「安定した正面注視」と判断し，FOCUSED を出力する.図 11(b)は DISTRACTED（Neutral: OK） と判定された例であり，頭部回転に伴って Yaw/Pitch が変化し，相対角が正面から大きく逸脱している.加えて，画面下部の HoldScore が上昇しており，一定方向への向き（よそ見）が継続したことが示される.これにより，瞬間的な視線移動ではなく「継続的に注視対象から外れている状態」として DISTRACTED に分類される.図 11(c)は閉眼時の例である.Iris が NONE となり虹彩情報が取得できず，EAR も低値を示して閉眼に相当する特徴が現れている.本システムでは，このように視線推定が不可能または信頼できないフレームを NG として扱い，閉眼が一定時間継続した場合に DISTRACTED へ遷移する.図 11(d)は DROWSY（Neutral: OK） と判定された例であり，視線の情報が取得できていても，脈拍が平常時基準から 20%低下するなど覚醒度低下の兆候が観測される場合に，眠気・ぼんやり状態として DROWSY を出力する.

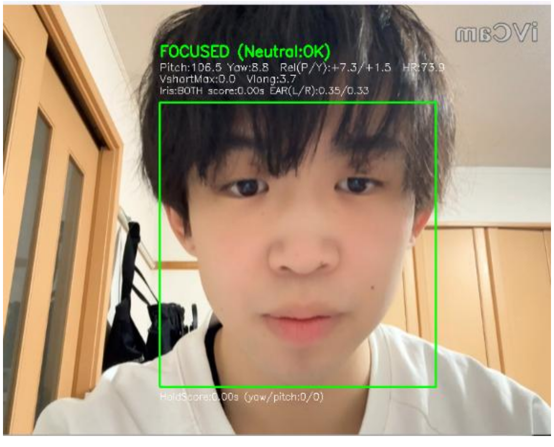


図 11(a)集中状態

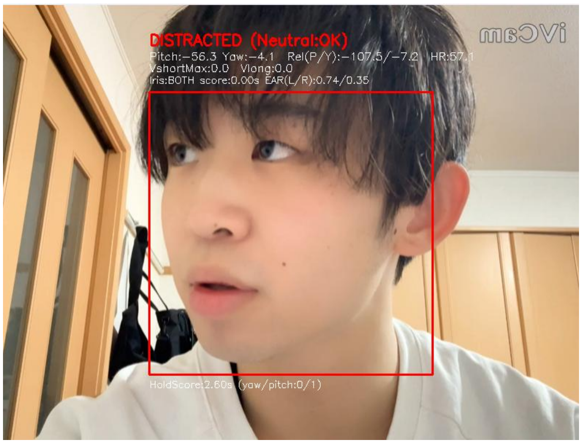
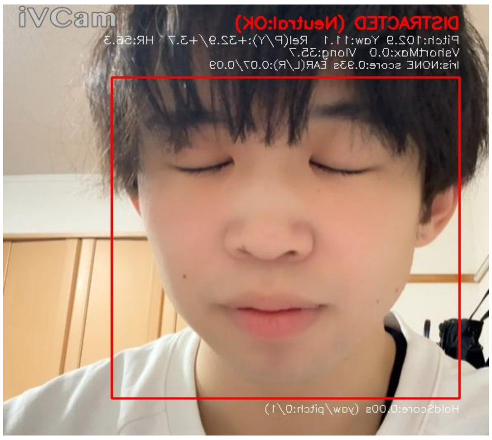


図 11(b)非集中状態(よそ見)



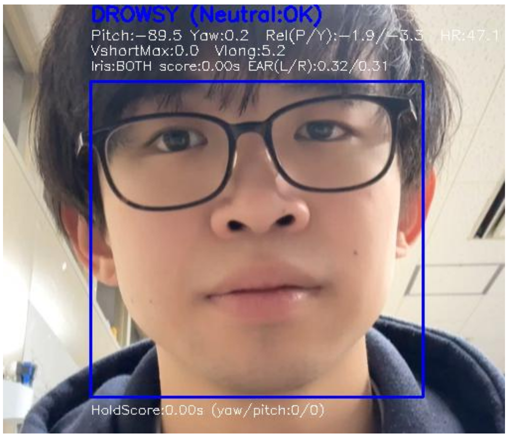


図 11(c)非集中状態(目をつむる)　　　　図 11(d)眠気,ボーっとしている

また，図 12 は処理終了後に出力される時系列ログ例である.判定結果はフレーム単位ではなく区間（例：0:00–0:02）としてまとめて出力され，各区間の継続時間（秒）が付与される.これにより，どの時間帯に集中していたか，どの時間帯に注意が逸れていたかを客観的なタイムラインとして把握できる.

```
👤 User ID: 1
------------------------------
[00:00 - 00:02] FOCUSED    (2.2s)
[00:02 - 00:06] DISTRACTED (4.2s)
[00:06 - 00:09] FOCUSED    (3.0s)
[00:09 - 00:11] DISTRACTED (2.2s)
[00:11 - 00:13] FOCUSED    (1.9s)
[00:13 - 00:33] DISTRACTED (20.0s)
[00:33 - 00:38] FOCUSED    (5.2s)
[00:38 - 00:45] DISTRACTED (7.2s)
[00:45 - 00:47] FOCUSED    (1.8s)
[00:47 - 01:03] DISTRACTED (15.4s)
[01:03 - 01:07] FOCUSED    (4.3s)
[01:07 - 01:23] DISTRACTED (15.6s)
[01:23 - 02:32] FOCUSED    (69.8s)
[02:32 - 02:37] DISTRACTED (4.8s)
[02:37 - 02:47] FOCUSED    (9.4s)
[02:47 - 02:52] DISTRACTED (5.8s)
[02:52 - 03:01] FOCUSED    (8.9s)
========================================
```

図 12 時系列判定ログ

4.2 考察

上記の結果より，本システムは視線推定値（Yaw/Pitch）とその変化量（視線速度），および目の状態（Iris，EAR）を統合することで，作業者の状態をリアルタイムに識別できる可能性を示した.特に，FOCUSED と DISTRACTED の判定が画面上の色変化として即座に可視化される点は，使用者にとって理解しやすく，セルフマネジメント支援として有効であると考えられる.また，本システムの重要な点として，視線速度のみに依存せず，複数の根拠を組み合わせて判定を行っている点が挙げられる.例えば，図 10(b)のような「横を向く」行為では，相対角（Rel）が大きくなり，かつ HoldScore が上昇することで，注視対象から外れた状態を継続として捉えられている.これにより，単に視線が一瞬動いた場合と，注視対象から離脱している場合を区別し，誤判定を抑える効果が期待できる.一方で，課題も確認された.図 10(c)のように閉眼時には虹彩情報が取得できず，視線推定が不安定となる.閉眼が必ずしも注意散漫を意味するとは限らず，瞬目や短時間の目閉じが頻繁に起こる状況では DISTRACTED が過剰に出力される可能性がある.この点については，閉眼を状態推定に反映するまでの継続時間を調整する，あるいは閉眼は一時的に状態判定を保留する等の設計改善が必要である.脈拍（HR）については，画面上に値が表示されていることから，顔映像からの非接触推定が実時間で行えていることが確認できる.しかし，脈拍は短時間で急変しにくい指標であり，視線よりも反応が遅れる.そのため，本システムにおける脈拍の役割は「瞬間状態の決定」ではなく，「集中低下が継続した場合の補助根拠」または「覚醒度低下の兆候検出」として位置づけるのが適切であると考えられる.

第 5 章 まとめと今後の展望

5.1 まとめ

本研究では，作業者の状態を非接触で推定することを目的として，顔映像から視線情報と脈拍情報を同時に取得し，それらを統合して状態を評価するシステムを構築した.近年，オンライン学習や PC 作業の長時間化により, 集中状態の低下や疲労の蓄積が作業効率に影響を及ぼすことが課題となっている.しかし，集中状態は主観的な判断に依存しやすく，定量的な把握が困難である.また，生体センサーを装着する方法は高精度である一方，装着負担が発生し，日常的な利用には適さない場合がある.そこで本研究では，映像のみを入力として状態推定を行う枠組みを提案した．まず，MediaPipe を用いて顔領域を高速に検出し，視線推定に必要な顔画像を安定して取得する基盤を整備した.次に，視線推定には L2CS-Net を採用し，Yaw 角および Pitch 角として視線方向を推定した.さらに，連続フレーム間の変化量から視線速度（velocity）を算出し，眼球運動の速さを定量化することで，集中状態の評価に利用可能な指標を得た.加えて，顔映像に含まれる微細な色情報変化に着目し，Eulerian Video Magnification（EVM）の原理を応用することで非接触型の脈拍推定を行っ

た.実装面では，画質の安定性を確保するため iVCam を用いてスマートフォンカメラ映像を入力とし，OpenCV によりフレーム列として取得した.そして，顔検出・視線推定・脈拍推定を統合し，リアルタイムに推定結果を取得・表示できることを確認した.これにより，装着型センサーを用いずに，日常環境に導入可能な状態推定システムの実現可能性を示した.

5.2 今後の展望

今後の展望として，第一に推定精度および安定性の向上が挙げられる.視線推定では，頭部運動や照明条件により推定角度に揺れが生じるため，顔領域検出の安定化や角度系列の平滑化をより最適化する必要がある.また，視線速度（velocity）の算出においては，処理フレームレートの変動や外れ値の混入が指標を不安定にする可能性があるため，外れ値除外や窓幅設計の妥当性を検討することが重要である.第二に，状態判定ロジックの確立が必要である.現段階では視線方向・視線速度・脈拍といった指標の取得が可能となったが，それらをどのように統合して「集中」「注意散漫」「眠気」等へ分類するかは今後の課題である.特に集中状態は個人差が大きく，単一の固定閾値では十分な汎用性を得られない可能性がある.したがって，被験者ごとの基準値の導入，あるいは機械学習による分類器の適用など，個人差を吸収する方法を検討する必要がある.

謝辞

本研究及び論文の作成にあたり，研究の着想や論文執筆等，多くのご指導，ご助言をいただきました静岡大学工学部の岡部誠准教授に心から感謝申し上げます.

参考文献

[1] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, Matthias Grundmann, “MediaPipe: A Framework for Building Perception Pipelines,” In arXiv 2019.
[2] Valentin Bazarevsky, Yury Kartynnik, Andrey Vakunov, Karthik Raveendran, Matthias Grundmann, “BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs,” In arXiv 2019.
[3] Tadas Baltrušaitis, Amir Zadeh, Yao Chong Lim, Louis-Philippe Morency, “OpenFace 2.0: Facial Behavior Analysis Toolkit,” In IEEE 2018.
[4] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, Antonio Torralba, “Eye Tracking for Everyone,” In CVPR, 2016.
[5] Xucong Zhang, Yusuke Sugano, Mario Fritz, Andreas Bulling, “Appearance-Based Gaze Estimation in the Wild,” In CVPR, 2015.
[6] Petr Kellnhofer, Adrià Recasens, Simon Stent, Wojciech Matusik, Antonio Torralba, “Gaze360: Physically Unconstrained Gaze Estimation in the Wild,” In ICCV, 2019.
[7] Xucong Zhang, Seonwook Park, Thabo Beeler, Derek Bradley, Siyu Tang, Otmar Hilliges, “ETH-XGaze: A Large Scale Dataset for Gaze Estimation under Extreme Head

Pose and Gaze Variation," In ECCV, 2020.
[8] Tobias Fischer, Hyung Jin Chang, Yiannis Demiris, "RT-GENE: Real-Time Eye Gaze Estimation in Natural Environments," In ECCV, 2018.
[9] Ahmed A.Abdelrahman, Thorsten Hempel, Aly Khalifa, Ayoub Al-Hamadi, "L2CS-Net: Fine-Grained Gaze Estimation in Unconstrained Environments," In ICIP 2022.
[10] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, William T. Freeman, "Eulerian Video Magnification for Revealing Subtle Changes in the World," In SIGGRAPH 2012.
[11] Neal Wadhwa, Michael Rubinstein, Frédo Durand, William T. Freeman, "Eulerian Video Magnification and Analysis," In Communications of the ACM 2017.
[12] Neal Wadhwa, Michael Rubinstein, Frédo Durand, William T. Freeman, "Phase-based Video Motion Processing," In SIGGRAPH 2013.
[13] Mahmoud Elgharib, Hossam (H.) E. Ammar, Michal (Michael) Kowalski, Piotr (Peter) Didyk, Hans-Peter Seidel, Christian Theobalt, "Video Magnification in Presence of Large Motions," In CVPR 2015.
[14] Weixuan Chen, Daniel McDuff, "DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks," In ECCV 2018.
[15] Xin Liu, Joshua Fromm, Shwetak Patel, Daniel McDuf "Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement,"In NeurIPS, 2020.
[16] Zitong Yu, Xiaobai Li, Guoying Zhao, "Remote Photoplethysmograph Signal Measurement from Facial Videos Using Spatio-Temporal Networks," In BMVC, 2019.
[17] Zitong Yu, Xiaobai Li, Guoying Zhao, "PhysFormer: Facial Video-based Physiological Measurement with Temporal Difference Transformer,"In CVPR, 2022.
[18] Xin Liu, Brian Hill, Ziheng Jiang, Shwetak Patel1, and Daniel McDuff, "EfficientPhys: Enabling Simple, Fast and Accurate Camera-Based Cardiac Measurement," In WACV, 2023.
[19] E. M. Nowara, T. K. Marks, H. Mansour, A. Veeraraghavan, "SparsePPG: Towards Driver Monitoring Using Camera-Based Vital Signs Estimation," In CVPR Workshops 2018.
[20] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham J. Mysore, Frédo Durand, William T. Freeman, "The Visual Microphone: Passive Recovery of Sound from Video," In SIGGRAPH 2014.