

美術作品の自由鑑賞時の視線計測データに対する特徴マッチングを用いた簡便な解析ツールの開発

楊 詩韻[†] 川畑 秀明^{†‡}

[†]慶應義塾大学大学院社会学研究科 〒108-8345 東京都港区三田 2-15-45

[‡]東京藝術大学社会連携センター 〒110-0007 東京都台東区上野公園 12-8

E-mail: ysyrah@keio.jp, kawabata@flet.keio.ac.jp

あらまし 視線計測は、様々な場所や状況に応じた装置や解析手法が必要となる。特に頭部を固定することない自由な状況での視線計測では、確立された方法論が乏しい。本研究では、特に美術鑑賞のような生態学的妥当性の高い自由鑑賞において、簡便かつ安価な計測システムを用いて視線を計測し、その際の視線計測データの解析や可視化のために特徴マッチングの手法を用いたツールを開発した。この手法は、被験者中心の座標系から観察する画像中心の座標系に変換し、計測データを精度良くかつ簡便に利用するためのツールとなっている。また、本研究では、絵画パネルを鑑賞している時の視線計測をもとに、この方法の妥当性についてテストした。

キーワード 美術鑑賞, 視線計測, 特徴マッチング, 解析ツール

1. 背景・目的

画像等を観察している時の目の動きを検出し分析することで、対象に対する注意や意思決定等の人間の情報処理過程を定量的に検討することができる。この方法は心理学、医学、神経科学、工学などの分野に広く応用され、計測する場所や状況に応じて、メガネ型、ヘッドマウント型、デスクトップ型など様々な視線追跡装置が開発されてきた[1-2]。

近年の実験心理学における研究では、デスクトップ型の計測装置を使って、頭や目の位置を固定して、モニター上の画像に対する視線の分布や推移を、視覚探索、読書、画像観察などの課題において、様々な認知活動が検討されてきた。また最近では、メガネ型視線計測装置を用いて、観察者が自由に動いていても視線データを測定することが可能になり、生態学的妥当性の高い状況において、実際の美術館での絵画鑑賞のような自由度の高い計測がなされるようになってきた[3-8]。実際の美術館や博物館での展示における鑑賞のあり方を実験的に検討することは、単に高い生態学的妥当性だけでなく、鑑賞における展示方法の影響について検討できるなど[6]の利点もある。

美術館や日常場面での美術鑑賞における視線計測を行う場合、実験そのものもできる限り日常的なセッティングで行う必要がある。その点では、メガネ型の装置を用いた視線計測は、比較的安価かつ簡便に研究環境を構築できる。メガネ型の視線計測器としては、例えば、Tobii 社の Pupil Pro Glasses 3 (<https://www.tobii.com/products/eyetrackers/wearables/tobii-pro-glasses-3>) や Pupil Labs 社の Pupil invisible などが挙げられる ([https://pupil-](https://pupil-labs.com/products/invisible)

[labs.com/products/invisible](https://pupil-labs.com/products/invisible))。これらの装置は観察者がメガネをかけるように装着でき、搭載されたフロントカメラにより観察者の視野に基づく動画を記録するとともに、角膜反射等により目（視線）の動きを記録する。また、キャリブレーションも簡便であることが多い。ただし、それらの装置では、メーカー独自のソフトウェアを用いて解析するか、ソフトウェア上から視線の時空間のデータを外部出力し、計算処理をして解析することが必要となる。

しかし、メガネ型装置を用いた視線計測では、観察者の動き（歩行等の移動、頭部や身体の動き等）により、搭載されたカメラによる録画も常に移動し、視野内の目標刺激の位置も変化することとなる。そのため、メガネ型視線計測装置を用いて、生態学的妥当性の高い自由鑑賞における視線計測データを解析するためには、視野の中に移動している目標刺激の位置を確定し、観察対象（例えば、絵画や部屋等）との対応関係の計算が必要となる。例えば、Pupil Invisible を使用する際には、データクラウド Pupil Cloud 上で“reference image mapper”という Pupil invisible で得られた視線計測データを使い、アップロードされた画像にヒートマップを作成する機能があるが、精度や成功率が低く、処理したデータ処理のアルゴリズムも明確でない。

本研究では、コンピュータビジョンの手法を用いて、観察者中心の座標系で取られた視線計測データを解析し、実験刺激画像中心の座標系に変換し、ヒートマップを生成する解析ツールを開発した。

2. 方法

2.1 装置

視線計測には Pupil Labs 社のメガネ型視線計測装置 Pupil invisible 2 台を用いた。

2.2 刺激

実験 1 において、実験刺激画像として、高解像度の静物画 2 点、を用いた。1 つは Edouard Manet “Still Life: Fruits on a Table” (1864 年作, オルセー美術館蔵; Figure 1 左), もう 1 つは Paul Cezanne “Still Life with Red Onions” (1898 年作, オルセー美術館蔵; Figure 1 右) であった。これらの絵画画像は高精細インクジェットプリンター (Epson SC-PX1V) を用いてマットフォト用紙に 27cm×45cm (Figure 1 左), 30 cm×37 cm (Figure 1 右) の大きさで印刷し, 33cm×48cm の 5 ミリ厚のスチレンボードの中央部に貼りつけた。



Figure 1 実験 1 において視線計測に用いた刺激画像。
左 : Edouard Manet “Still Life: Fruits on a Table”。
右 : Paul Cezanne “Still Life with Red Onions”。



Figure 2 実験 2 において視線計測に用いた刺激画像。
左 : Pierre-Henri de Valenciennes “Classical Landscape with Figures and Sculpture”。右 : Edgar Degas “A Cotton Office in New Orleans”。

実験 2 において、実験刺激画像として高解像度の風景画 1 点と風俗画 1 点の合計 2 点を用いた。風景画は Pierre-Henri de Valenciennes “Classical Landscape with Figures and Sculpture” (1788 年作, J. Paul Getty Museum 蔵; Figure 2 左), 風俗画は Edgar Degas “A Cotton Office in New Orleans” (1873 年作, Musee des Beaux-Arts de Pau 蔵; Figure 2 右) であった。これらの絵画画像は高精細大

判インクジェットプリンター (Epson SC-P9550) を用いてキャンバス用紙に原画のサイズで印刷し, パネルフレームの中央部に装着した。風景画のサイズは 29 × 41 cm, 風俗画のサイズは 73 x 92 cm であった。

2.3 データ

実験 1 において、実験参加者 12 人が Pupil invisible を装着し、実験刺激画像を印刷し、絵画貼付用ボードに貼った物を目標刺激として観察した。Pupil Cloud により得られた視線計測データは計 24 であった。しかし、そのうち装置の不調により 9 のデータが欠損しており、計 15 であった。

実験 2 において、実験参加者 4 人が Pupil invisible を装着し、実験刺激画像を印刷し、額装した絵画を目標刺激として観察した。Pupil Cloud により得られた視線計測データは計 8 であった。

実験 1 と実験 2 を合わせて、合計 23 のデータを分析の対象として用いた。視線計測データには装置のフロントカメラで撮影された動画及び観察者中心の注視点座標情報が含まれた[9]。Figure 4 はフロントカメラで撮影された動画に赤丸で表示された注意位置をスーパーインポーズし、スナップショット化したものである。



Figure 3 実験 2 において実験参加者が観察している姿



Figure 4 計測装置に取り付けられたカメラによって録画された画像と注視点とを重ね合わせたもの

2.4 解析手法

視野全体をできるだけ含めるために、Pupil invisible では魚眼レンズを使ったビデオを録画し、魚眼補正後、画面周辺が切り取られる。しかし近距離で刺激を観察する場合、注視の対象物となる目標刺激全体は視野となるビデオの画面に入らない可能性があるため、目標刺激をできるだけ画面に入れるために本研究では魚眼補正をしないまま、次の解析に進んだ。以下は、視線解析のための基本的なステップを4段階で示す。

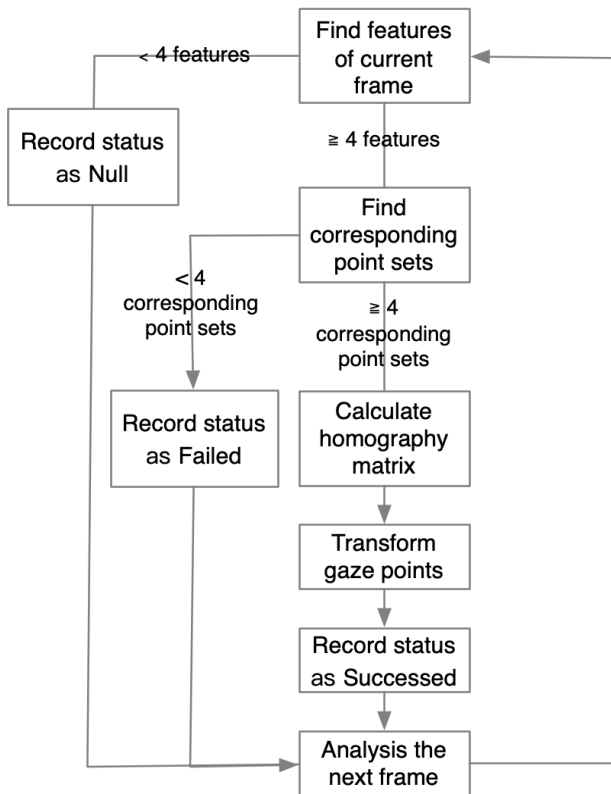


Figure 5 解析の流れ

(1) 特徴抽出

観察の対象となる目標刺激は絵画でありオブジェクトが含まれているため、Mask R CNN[10]のようなオブジェクト認識アルゴリズムだけでは絵画そのものを識別しにくい。その一方で、SIFT (Scale-Invariant Feature Transform) アルゴリズム[11-12]は局所的な注目点ベースのアルゴリズムであり画像のサイズや回転に影響されないだけでなく、照明やノイズなどの影響にも強いという特徴を持ち、優れた頑健性を示し、より高い精度のマッチング結果を得るために使用できる[13]。本研究では SIFT アルゴリズムを用いて実験刺激画像の特徴及びビデオのフレームごとの特徴を抽出した。

(2) 特徴マッチング

FLANN (Fast Library for Approximate Nearest Neighbors) は高速最近傍点探索ライブラリで、近傍探索に最適と思われるアルゴリズム集が含まれている[14]。FLANN 中の KNN (K- Nearest Neighbor) アルゴリズムはトレーニングセットとテストデータとのユークリッド距離を求め、上位 K 個のデータの距離に対するテストデータの割合に応じて上位 K 個のデータのカテゴリを組み合わせる計算し、テストデータを分類する方法である。前述の SIFT では、各特徴点は 128 次元の特徴を持ち、テンプレートの 128 次元特徴との距離に基づいて計算され、K は 1 とされる。つまり、距離が最も小さい点に対応するマッチング点である可能性が高い。SIFT と KNN を基づく方法が正しいマッチングを保持すると同時に、90%以上のミスマッチを排除し、正しいマッチング率を大幅に向上させることができる[15]。そのため、本研究では KNN アルゴリズムを用いて二つの画面から抽出した特徴の間に、特徴マッチングを行った。

(3) 行列計算

本研究では RANSAC (Random Sample Consensus) で複数の特徴の間の最適なマッピングの変換行列を算出した。RANSAC とは、総数 N のサンプル点の集合から、K 個のサンプル点をランダムに抽出し、モデルを当てはめる。他の点からフィットしたモデルまでの距離を計算し、ある閾値以下の点を内部点と見なし、内部点の数をカウントし、最も内点数の多いモデルを選択するアルゴリズムである[16]。

(4) 座標変換

RANSAC により求めた行列は 3×3 の行列で、Normalizer があるので、次の算式により、フレームごとに視線計測データの座標を実験刺激画像において対応した座標に変換した。

$$(\hat{y}_0, \hat{y}_1, Z) = (x_0, x, 1) \times \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix}_{3 \times 3} \quad (1)$$

$$(y_0, y_1) = \left(\frac{\hat{y}_0}{Z}, \frac{\hat{y}_1}{Z} \right) \quad (2)$$

2.5 可視化

目標刺激と対応したエリアをフレームごとにプロットし、マッチングの結果を Figure 7 や Figure 8 の通りに可視化した。Figure 7 の左の上を例として、左は実験刺激画像に水色の丸で表示された座標系を変換した注意位置をスーパーインポーズしたもので、右はフロントカメラで撮影された動画のフレームに水色の丸で表示された注意位置をスーパーインポーズしもので、左の実験刺激画像と右のビデオのフレームの間の緑の線はマッチングした特徴の対の対応関係を示した

もので、右のビデオのフレームにおける目標刺激の周りの青い線はマッチングした特徴を囲む輪郭である。

また、GUI (graphical user interface) により、実験刺激データを選択することができ、変換した座標の参照を CSV のフォーマットでアウトプットしてできるようにした。また、Sun ら[17]に従い、実験刺激画像における変換した座標により、“pyHeatMap” (<https://github.com/oldj/pyheatmap>)という Python のライブラリを用いて、ヒートマップの生成もできる。

2.6 妥当性テスト

試行ごとに対して、Baldi ら[18]に従い、Accuracy を計算した。Accuracy の計算は算式 (3) の通りであり、TP (True Positive), TN (True Negative), TP (True Positive), TP (True Positive) の判断基準は Figure 6 のとおりであった。Pupil Invisible では、装置が記録を開始した時、最初の 1~2 秒間は真っ暗の画面で、Null となった。状態が Succussed と Failed のフレームに対して、フレームの画面中、目標刺激があるか、マッチングした特徴を囲む輪郭は正確に目標刺激を囲むかを手動で確認した。そのため、実際とは若干のずれがあった。

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Stimuli	Frame status	Frame content	Contour check	Conclusion
A	Succeeded	Containing A	Correct	True Positive
		Not containing A	-	False Positive
	Failed	Containing A	-	False Negative
		Not containing A	-	True Negative
	Null	None	-	True Negative

Table 1 妥当性テストの基準。Frame status : フレームの解析が成功したかを示す指標。Frame content : フレームの画面中、目標刺激があるかを示す指標。Contour check : マッチングした特徴を囲む輪郭は正確に目標刺激を囲むかを示す指標。Conclusion : 判断の結果

3. 結果

本研究では、実験 1 と実験 2 を合わせて、4 の実験刺激画像に対して、合計 23 のデータを分析の対象として用いた。データを解析した結果、刺激ごとの実験刺激画像における変換した視線計測データの座標により生成したヒートマップは Figure 6 の示した通りであった。左は個人レベルの例で、右は平均した全体レ

ベルのヒートマップであった。1 行目の“Still Life: Fruits on a Table”における注意位置は箱の中の果物に集まり、右のガラスにもあった。2 行目の“Still Life with Red Onions”における注意位置はテーブルの上の果物と瓶に集まった。3 行目の“Classical Landscape with Figures and Sculpture”における注意位置は左下の二人、及び画面の中心の遠くに見える森の中の城に集まった。4 行目の“A Cotton Office in New Orleans”における注意位置は画面の中心の人々の顔のところであった。

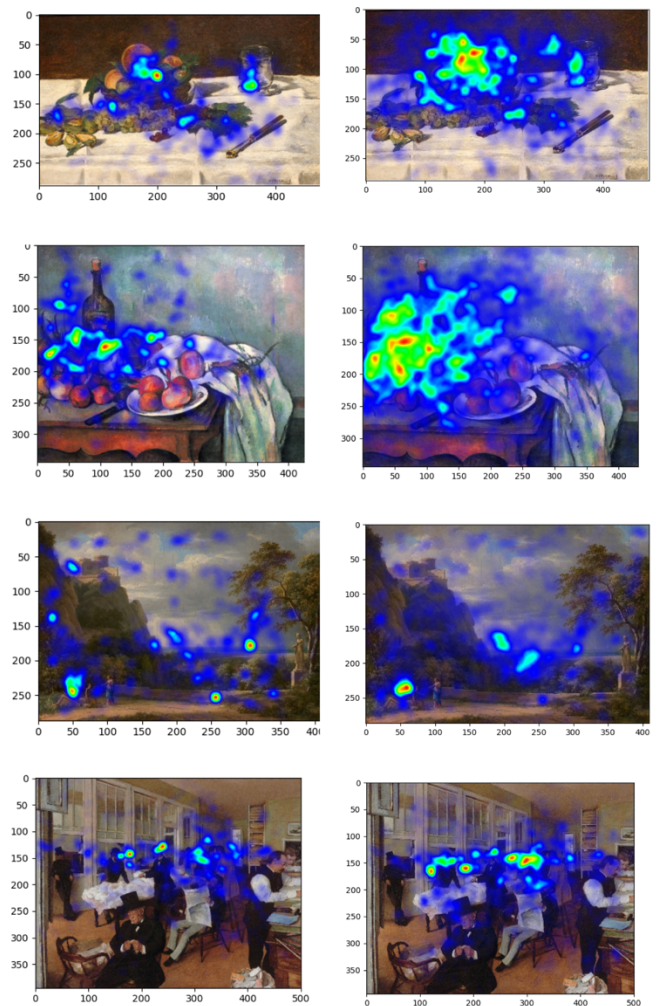


Figure 6 刺激ごとのヒートマップ。左：個人レベルのヒートマップの例。右：全体レベルのヒートマップ

妥当性テストの結果、平均的には、Accuracy は 99.15%に達し、標準偏差は 0.03 であった。刺激ごとのデータ数、平均 Accuracy と標準偏差、平均フレーム数 (即ち TP+TN+FP+FN) , 平均 TP 数、平均 TN 数は Table 2 の通りであった。例として、Figure 7 と Figure 8 の示した通りであった。Figure 7 の左上、左中、左下、右中、右下が TP で、右上が FP であった。Figure 8 の左中、左下、右下が TP で、左上、右上、

Stimuli	Trial	Average			
		Accuracy(SD)	Frames	True Positive	True Negative
Still Life: Fruits on a Table	6	99.96%(<0.01)	1904.17	1875.17	28.17
Still Life with Red Onions	9	99.97%(<0.01)	2081.11	2063.00	17.56
Classical Landscape with Figures and Sculpture	4	95.35%(0.06)	1865.25	1781.00	73.50
A Cotton Office in New Orleans	4	99.79%(<0.01)	1930.00	1876.75	49.25
Total	23	99.15%(0.03)	1971.13	1932.57	35.57

Table 2 刺激ごとのデータ数，平均 Accuracy と標準偏差，平均フレーム数，平均 TP 数，平均 TN 数

Stimuli	Trial	Average		
		Processing speed(SD)	Video duration(min)	Runtime(min)
Still Life: Fruits on a Table	6	10.08(0.85)	1.16	11.65
Still Life with Red Onions	9	10.18(1.24)	1.18	12.05
Classical Landscape with Figures and Sculpture	4	8.50(0.40)	1.18	10.01
A Cotton Office in New Orleans	4	10.66(0.96)	1.08	11.50
Total	23	8.94(1.21)	1.16	11.49

Table 3 刺激ごとのデータ数，平均解析速度と標準偏差，平均映像時間，平均解析時間

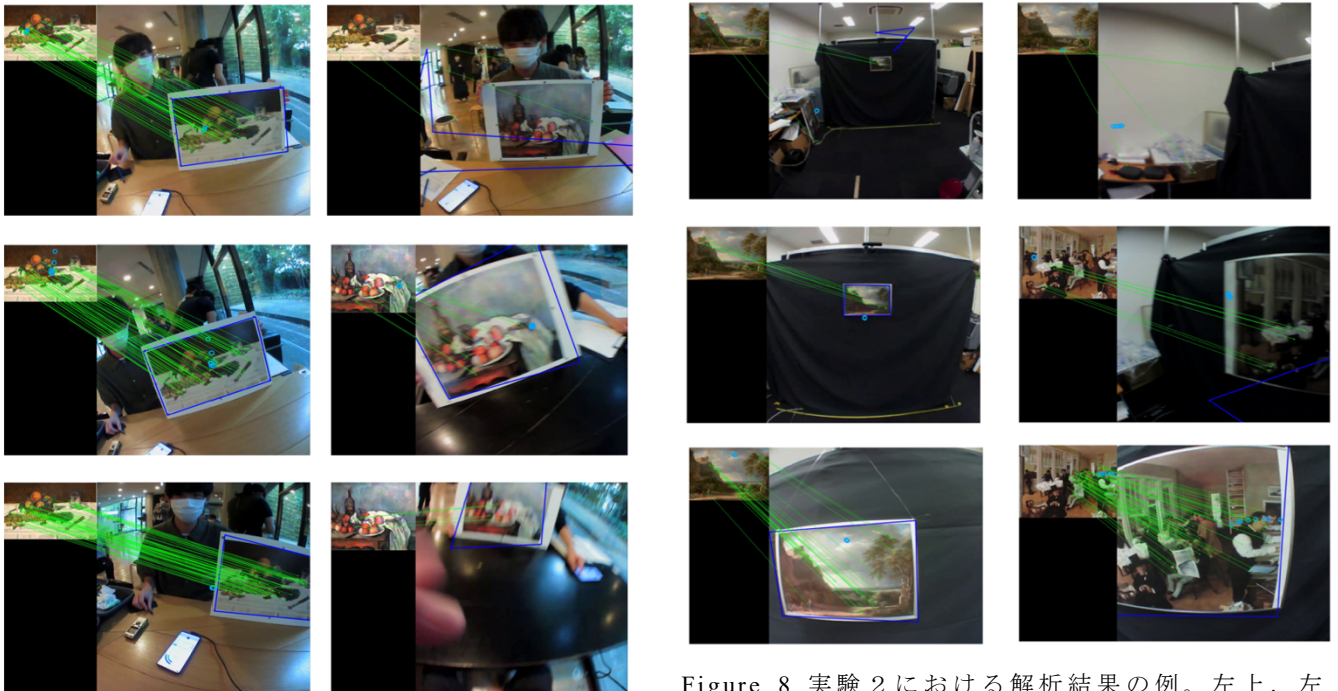


Figure 7 実験 1 における解析結果の例。左上，左中，左下：“Still Life: Fruits on a Table”を含むフレームにおける解析結果。右上：“Still Life: Fruits on a Table”を含まないフレームにおける解析結果。右中，右下：“Still Life with Red Onions”を含むフレームにおける解析結果

右中が FP であった。4 つの実験刺激画像の内，“Still Life with Red Onions”の Accuracy は一番低く，平均 95.35%で，標準偏差が 0.06 であった。他の 3 つの実験刺激画像は平均 Accuracy が 99%以上で，標準偏差が 0.01 未満であった。なお，解析は CPU が 3.60GHz，GPU が 11264Mib のコンピュータで行なった。解

Figure 8 実験 2 における解析結果の例。左上，左中，左下：“Classical Landscape with Figures and Sculpture”を含むフレームにおける解析結果。右上：“Classical Landscape with Figures and Sculpture”を含まないフレームにおける解析結果。右中，右下：“A Cotton Office in New Orleans”を含むフレームにおける解析結果

析の速度は Table 3 の通りで，1 分間のビデオが平均 8.94 分間かかった。

4. 考察

本研究では，メガネ型視線追跡装置により計測されたデータを，精度や成功率が高く解析するうえ，処

理したデータの結果を可視化するために、コンピュータビジョンの手法を用いて、観察者中心の座標系で取られた視線計測データを解析し、実験刺激画像中心の座標系に変換し、ヒートマップを生成する解析ツールを開発した。

実験の結果、個人レベルにも全体レベルにもヒートマップを高い成功率で生成された。また、平均 Accuracy が 99%以上のことから、本研究で使った解析手法は精度が高いことがわかった。特に、実験 1 においては、被験者と目標刺激との観測距離が比較的一定であったため、目標刺激が斜めになった場合や一部が欠落した場合でも良好に解析できた。実験 2 においては観察画像が大きい場合には遠距離でもマッチングがうまく行われるが、小さい場合には観察からの距離が遠くなるとフロントカメラで撮影された目標刺激が鮮明でなくなり、マッチングための特徴を十分に捉えられなくなり、特徴解析が失敗することが多くなった。また、画面の動きが速い場合や対象刺激の欠落部分が多い場合、この解析手法の妥当性が落ちる。従って、このツールを使用する際には、参加者があまり速く動きすぎないことと、目標刺激ができるだけ視野の大部分を占めるようにすることが制限である。

ヒートマップについて、本研究では、最もシンプルに注意位置の座標のみ用いてヒートマップを作成し、まばたきの状態や注視時間など[19-20]は扱わなかったため、今後、解析ツールを改善する必要がある。

謝辞：本研究は、JST 共創の場形成支援プログラム「『共生社会』をつくるアートコミュニケーション共創拠点」(JPMJPF2015)の支援を受けた。

文 献

- [1] M. L. Mele, and S. Federici, Gaze and eye-tracking solutions for psychological research. *Cognitive processing*, vol.13, pp.261-265, 2012.
- [2] B. T. Carter, and S. G. Luke, Best practices in eye tracking research. *International Journal of Psychophysiology*, vol.155, pp.49-62, 2020.
- [3] B. Balbi, F. Protti, and R. Montanari, Driven by Caravaggio through his painting. *Proc. COGNITIVE*, pp.72-76, 2016.
- [4] S. Ashrafi, and M. Garbutt, Looking at looking: two applications of a hermeneutic phenomenological analysis of head-mounted eye-tracking data in gallery visitation. *Australian and New Zealand Journal of Art*, vol.17, no. 2, pp.221-237, 2017.
- [5] T., Santini, H. Brinkmann, L. Reitstätter, H. Leder, R. Rosenberg, W. Rosenstiel, and E. Kasneci, The art of pervasive eye tracking: Unconstrained eye tracking in the Austrian Gallery Belvedere. *Proc. the 7th workshop on pervasive eye tracking and mobile eye-based interaction*, pp.1-8, June 2018.
- [6] L. Reitstätter, H. Brinkmann, T. Santini, E. Specker, Z. F. Dare, Bakondi, ... and R. Rosenberg, The display makes a difference: A mobile eye tracking study on the perception of art before and after a museum's rearrangement. *Journal of Eye Movement Research*, vol.13, no.2, June 2020.
- [7] L. Reitstätter, K. Galter, and F. Bakondi, Looking to read: How visitors use exhibit labels in the art museum. *Visitor Studies*, vol.25,no.2, pp.127-150, 2022.
- [8] J. C. de Winter, D. Dodou, and W. Tabone, How do people distribute their attention while observing The Night Watch?. vol.51, no.11, pp.763-788, August 2022.
- [9] W. Fuhl, D. Weber, and E. Kasneci, Pistol: Pupil Invisible Supportive Tool to extract Pupil, Iris, Eye Opening, Eye Movements, Pupil and Iris Gaze Vector, and 2D as well as 3D Gaze. January 2022.
- [10] K. He, G. Gkioxari, P. Dollár, and R. Girshick, Mask r-cnn. *Proc. ICCV*, pp. 2961-2969, 2017.
- [11] D. G. Lowe, Object recognition from local scale-invariant features. *Proc. ICCV*, vol. 2, pp. 1150-1157, September 1999.
- [12] D. G. Lowe, Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, vol.60, pp.91-110, 2004.
- [13] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, ... and L. V. Gool, A comparison of affine region detectors. *ICCV*, vol.65, no.1, pp.43-72, 2005.
- [14] M. Muja, and D. Lowe, Flann-fast library for approximate nearest neighbors user manual. Computer Science Department, University of British Columbia, Vancouver, Canada, 5, 2009.
- [15] A. Yong, and Z. Hong, SIFT matching method based on K nearest neighbor support feature points. *ICSIP*, pp. 64-68, August 2016.
- [16] J. Matas, and O. Chum, Randomized RANSAC with sequential probability ratio test. *ICCV* vol.1 no.2, pp.1727-1732, October 2005.
- [17] M. Sun, W. Zou, N. Hu, J. Wang, and Z. Chi, Iterative brain tumor retrieval for MR images based on user's intention model. *Pattern Recognition*, vol.127, July 2022.
- [18] P., Baldi, S. Brunak, Y. Chauvin, C. A. Andersen, and H. Nielsen, Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, vol.16, no.5, pp.412-424, 2000.
- [19] D. Venugopal, J. Amudha, and C. Jyotsna, Developing an application using eye tracker. *RTEICT*, pp.1518-1522, May 2016.
- [20] Y. Cho, Rethinking eye-blink: assessing task difficulty through physiological representation of spontaneous blinking. *Proc. CHI conference on human factors in computing systems*, May 2021