

AR Memory Viewer: 過去写真の AR 提示による思い出参照体験

殿岡 柊也^{1,a)} 入山 太嗣¹ 小室 孝¹

概要: 本論文では、過去写真の AR 提示を行うことで過去シーンを再現する *AR Memory Viewer* を提案する。本システムでは、ユーザがモバイル端末のカメラを起動して現在シーンに向けてると、同じ場所を撮影した過去写真が現在シーンに位置合わせされ、モバイル端末の画面上に違和感なく AR 提示される。モバイル端末のカメラが取得した現在シーンに対し、見た目が大きく異なる過去写真の位置合わせを実現するため、深層学習による幾何構造を踏まえた特徴点マッチングを行う。この手法により、季節の異なる風景や、ペットがいた時の過去シーンの AR 提示を実現した。試作システムの評価実験を行い、*AR Memory Viewer* が新しい思い出の参照体験を提供することを確認した。

1. はじめに

思い出は、過去に経験した出来事や感情の記憶として人生を豊かにするものであり、なかでも過去写真は古い記憶を呼び起こす契機を提供する。モバイル端末の普及により、写真を撮影する機会は増加したが、多くの写真はフォルダ内に保存されたまま活用されていない [1]。そのため、過去写真からより豊かな体験や価値を引き出す方法が求められている。

過去写真を有効活用する一つの方法として、現在シーンと同じ場所を撮影した過去写真を AR 提示することで過去シーンを再現する手法がある [2]。この手法では、過去シーンを違和感なく AR 提示するために、現在シーンに対して過去写真を正確に位置合わせすることが求められる。この要求に対する一つのアプローチとして、GPS を活用する手法がある [3, 4]。この手法では、ユーザの現在位置情報と近いジオタグを持つ過去写真を選択し、それを AR 提示することで過去の街並みを再現することを可能にした。しかし、この手法では、ジオタグがあらかじめ付与された過去写真のみ利用可能であるという制約があり、屋内環境では GPS の精度が低下して利用困難な場合が多い。

別のアプローチとして、特徴点マッチングを用いる手法がある [5]。この手法では、過去写真に写るランドマークと現地に残存する同じランドマークを特徴点マッチングで検出し、過去写真が撮影された位置を推定することで正確な位置合わせを実現した。しかし、従来の特徴点マッチン

グ手法では、天候や時間帯の違いによってマッチング精度が低下し、正確な位置合わせができない場合がある。

この課題に対し、異なる気象条件下で同じ場所を撮影した複数の参照画像を事前に準備することで多様な環境下でのマッチングによる位置合わせを可能にする手法が提案されている [6]。しかし、この手法では事前に用意した気象条件や時間帯での利用に限られる。そのため、過去写真と現在シーンの見た目が大きく異なる場合でも安定したマッチングによる位置合わせを実現するための手法が依然として求められている。

本研究では、この問題に対し、深層学習による特徴点マッチングを活用することにより見た目が大きく異なる場合でも幾何構造が類似していれば過去写真を現在シーンに正確に位置合わせして AR 提示できる *AR Memory Viewer* を提案する。*AR Memory Viewer* は、観光地での利用だけでなく、個人の思い出の過去シーンを端末越しに見る新しい思い出の参照体験を提供する。試作システムの評価実験を行い、新しい思い出の参照体験として有効であることを確認した。

2. AR Memory Viewer

2.1 提案インターフェース

提案インターフェースを図 1 に示す。図 1a に示すように、ユーザはモバイル端末のカメラを現在シーンに向けて提案インターフェースを使用する。モバイル端末が取得する現在シーンが写るカメラ画像に、同じ場所を撮影した過去写真がリアルタイムで位置合わせされ、モバイル端末の画面に違和感なく AR 提示される。これにより、ユーザは

¹ 埼玉大学

^{a)} s.tonoka.644@ms.saitama-u.ac.jp

モバイル端末を左右に動かしながら端末越しに過去シーンを見ることができる。

ユーザが提案インターフェースを使用した際の視点を図 1b に示す。ユーザとモバイル端末の距離と、モバイル端末と現在シーンの距離を仮定し、モバイル端末に表示する画像を拡大することで画面内外がシームレスにつながるマジックレンズとして表示する。これにより、モバイル端末越しに過去シーンを見る新しい思い出の参照体験を提供する。

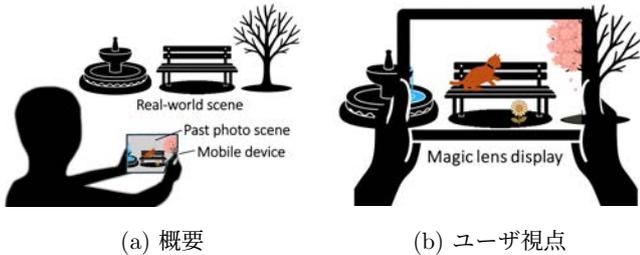


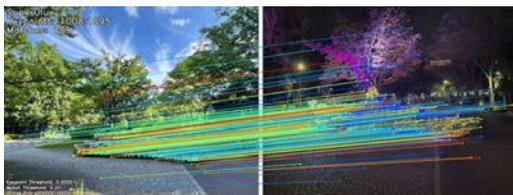
図 1: 提案インターフェース

2.2 深層学習による特徴点マッチング

提案インターフェースを実現するために、深層学習による特徴点マッチングを活用した処理を行う。特徴点抽出には深層学習により頑健な抽出が可能な SuperPoint [7]、マッチングには Attention 機構を利用したグラフニューラルネットワークを用いる SuperGlue [8] を使用した。従来のマッチング手法である AKAZE [9] と、SuperGlue によるマッチング手法のマッチング結果の比較を図 2 に示す。図 2a に示すように、AKAZE によるマッチングでは、同じ場所を撮影した写真でも見た目が大きく異なると正しくマッチングできていない。それに対し、SuperGlue によるマッチングは図 2b に示すように、見た目が大きく異なる場合でも幾何構造が類似していればマッチングが可能である。



(a) AKAZE によるマッチング



(b) SuperGlue によるマッチング

図 2: マッチング手法の比較

2.3 過去写真の検索と位置合わせ

前節で述べた深層学習による特徴点マッチングを用いて、ユーザが取得したカメラ画像に対し、同じ場所を撮影した過去写真の検索と位置合わせを行う。処理の概要を図 3 に示す。ユーザが取得したカメラ画像とフォルダ内のすべての過去写真で特徴点マッチングを行い、マッチングした特徴点の数が最も多い過去写真を同じ場所を撮影した過去写真とし、AR 提示する写真に選定する。選定した写真とマッチングした特徴点から射影変換を行い、カメラ画像に過去写真を位置合わせした画像を作成する。作成した画像に対しユーザ視点に合わせて拡大し、マジックレンズとしてユーザのモバイル端末に表示する。

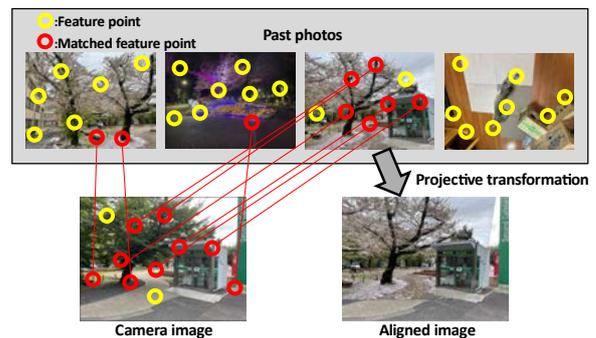


図 3: 過去写真の選定と位置合わせ

3. 実装

提案インターフェースを実現するための試作システムを実装した。本システムでは、SuperGlue を用いて過去写真の検索および位置合わせを行うが、SuperGlue は計算量が大きく、モバイル端末上でのリアルタイム実行は困難である。そこで、モバイル端末とデスクトップ PC を TCP 通信で接続し、計算負荷の高い処理をデスクトップ PC 側で行う構成とした。モバイル端末で取得した現在シーンのカメラ画像をデスクトップ PC に送信し、デスクトップ PC 上で SuperGlue によるマッチングを用いた画像検索と位置合わせを行う。位置合わせが完了した画像をモバイル端末に送信し、モバイル端末上では計算量の少ない AKAZE を利用した位置合わせを行うことで、カメラを動かした場合でも正確な位置合わせを実現する。

モバイル端末は Surface Pro 7 Model 1866 i7、デスクトップ PC は NVIDIA GeForce GTX 1070 GPU を搭載したものを使用した。SuperPoint によるフォルダ内の画像の特徴点抽出はシステムを起動する前に行う。SuperGlue によるマッチングを用いた画像検索にかかる時間は、フォルダ内の画像数が 1 枚増加するたびに 0.1 秒増加する。モバイル端末上での AR 提示は、約 5FPS で動作する。試作システムの動作結果を図 4 に示す。屋外では枯れ木状態の桜の木がある現在シーンに対し、桜が満開で咲いている季節

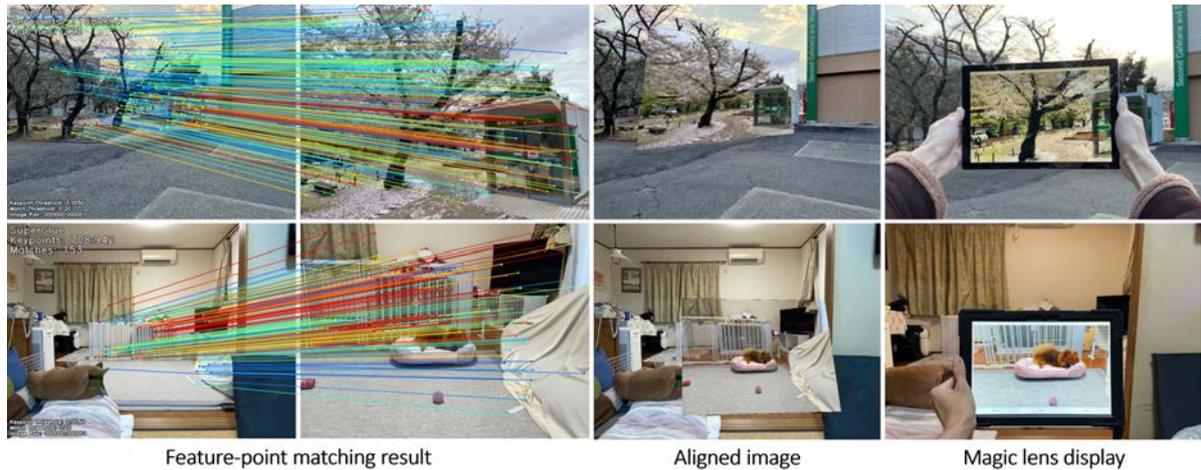


図 4: 試作システムの動作結果

の過去シーン、屋内ではペットがいない状態のリビングで、ペットがいた時の過去シーンを対象に動作を確認した。それぞれの現在シーンに対して同じ場所を撮影した過去写真の選定と正確な位置合わせが行われ、端末越しに過去シーンを見ることができることを確認した。

4. 評価実験

4.1 実験計画

提案インターフェースが新しい思い出の参照体験として有効であることを検証するために比較実験を行った。比較対象のベースライン手法は、現在シーンに類似する過去写真をフォルダから手動で選択し、画面中央に AR 提示する手法とした。フォルダ内の過去写真は、ベースライン手法で被験者が手動で選択する際の負担を考慮し、10 枚に設定した。内訳は、再現用の過去写真 1 枚に加え、写真共有サイト Flickr で、「indoor place」と「outdoor place」のキーワードで検索した写真の中から無作為に抽出した 9 枚で構成した。

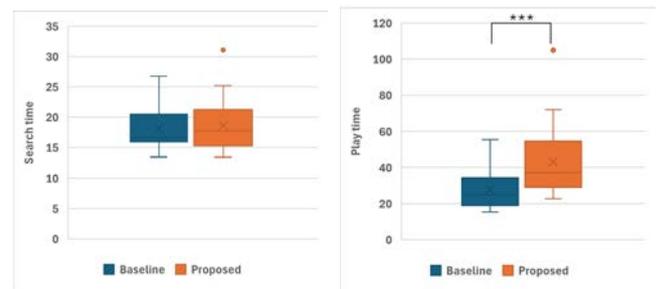
実験には本学の学生 20 名 (女性 5 名, 年齢 22.9 ± 1.74) が参加した。実験は 2025 年 7 月に実施した。本学の敷地内で撮影された過去写真 4 枚 (2025 年 6 月に屋内 2 か所で撮影された 2 枚, 2023 年 12 月および 2025 年 4 月に屋外 2 か所で撮影された 2 枚) を対象に、各写真が撮影された場所でいずれかの手法を体験してもらった。また、各場所において、いずれの手法も同数の被験者が体験するように割り当てた。各場所に被験者を誘導した後、提案インターフェースを体験する際にはカメラを向けるべき方向を指示した。また、それぞれのアプリケーションを十分楽しんだと感じたら終了するよう指示した。体験中には、アプリケーションを起動してから現在シーンと同じ場所を撮影した過去写真が選定されるまでの時間である画像検索の完了時間と、過去写真の AR 提示が開始してから、アプリケーションを終了するまでの時間であるアプリケーション使用

時間を計測した。すべての体験終了後、被験者には Lee らの研究 [3] で AR 体験の評価に用いられており、ユーザ体験を測定する指標として有効とされている GEQ [10] の設問内の「ゲーム」の語を「アプリケーション」変更したアンケートと、独自に作成した 3 つの設問に回答してもらった。

4.2 実験結果

4.2.1 計測時間

計測時間の実験結果を図 5 に示す。図 5a に示すように、画像検索の完了時間は、ベースライン手法で平均 18.1 秒 (± 3.65)、提案インターフェースで平均 18.56 秒 (± 4.31) となり、対応のある t 検定の結果、有意差は認められなかった。今回の実験では、フォルダ内の過去写真の枚数が 10 枚と少なかつたため、十分に評価できなかった可能性がある。今後、より多くの枚数を用いることで、より精度よく評価できると考えられる。図 5b に示すように、アプリケーション使用時間については、ベースライン手法が平均 27.82 秒 (± 11.81)、提案インターフェースが平均 43.13 秒 (± 19.89) となり、対応のある t 検定の結果、提案インターフェースの有意差が認められた。この結果から、提案インターフェースのほうがより長い時間アプリケーションを楽しんだことが示唆される。



(a) 画像検索の完了時間 (b) アプリケーション使用時間

図 5: 実験結果：計測時間 (***) は $p < .001$ を示す

4.2.2 GEQ のスコア

GEQ のコアモジュールを用いたアンケートの結果を図 6 に示す。GEQ のコアモジュールは、アプリケーション体験を有能感 (competence), 没入感 (sensory and imaginative immersion), フロー (flow), 緊張/ 不快感 (tension / annoyance), 挑戦 (challenge), 否定的感情 (negative affect) および肯定的感情 (positive affect) の 7 つの評価尺度に対して、数値が大きいほど当該設問に対する肯定度が高くなる 5 段階で評価する。GEQ において、それぞれの評価尺度のスコアは、その評価尺度に対応する項目の平均値である。対応のある t 検定の結果、有能感, 没入感, フロー, 肯定的感情, 否定的感情のスコアで有意差が認められた。この結果から、提案インターフェースが被験者に余計な負荷や難しさを与えることなく AR アプリケーションにおける体験の質を総合的に引き上げていることを示唆している。

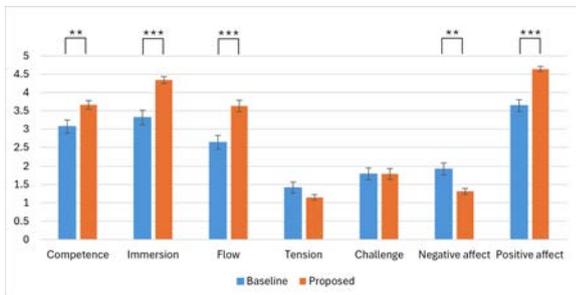


図 6: GEQ のスコア (** は $p < .01$, *** は $p < .001$, エラーバーは標準誤差を示す)

4.2.3 独自設問

独自設問のアンケート結果を図 7 に示す。被験者には、提案インターフェースに対する印象を評価するため、コンセプト理解、ユーザーの使用意欲、快適性を確認する 3 つの独自設問を用意した。それぞれの設問に対し、数値が大きいほど肯定的な回答を示す 1~7 の 7 段階評価で回答してもらった。

コンセプト理解は、提案インターフェースのコンセプトが被験者に伝わったかどうかを確認するための指標である。本指標に対する設問として、「モバイル端末越しに過去シーンを見ていると感じたか」を設定した。7 と 6 の回答が全体の 80 % を占めており、提案インターフェースのコンセプトが被験者に十分に伝わったことがうかがえる。

ユーザーの使用意欲は、提案インターフェースを使用して自分の思い出シーンを再現したいという被験者の使用意欲を評価する指標である。本指標に対する設問として、「このアプリケーションで自分の思い出シーンを再現したいと感じたか」を設定した。すべての回答が 7 または 6 となり、提案インターフェースが新しい思い出の参照体験として有効であることが示唆される。

快適性は、過去写真の選択および配置に関して、システ

ム使用時における快適さを評価する指標である。本指標に対する設問として、「写真の選択や配置が快適だったか」を設定した。すべての回答が 7, 6, 5 であり、写真選択の負担が軽減され、過去写真がカメラ画像に対して継続的に正確な位置に AR 提示されたことがうかがえる。

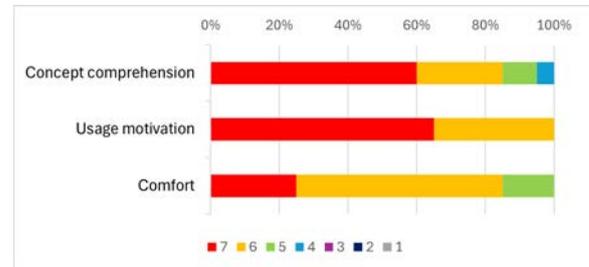


図 7: 独自設問の結果

5. まとめ

本研究では、深層学習による特徴点マッチングを活用し、個人の思い出の過去写真を正確な位置に AR 提示可能な *AR Memory Viewer* を提案した。*AR Memory Viewer* では、過去写真の外観が現在シーンと大きく異なる場合でも、幾何学的構造が類似していれば端末越しに過去シーンを AR 提示できる。評価実験の結果から、被験者にコンセプトが十分に伝わり、新しい思い出の参照方法として有効であることを確認した。

今後の課題として、ユーザのカメラ角度の変化に対応した正確な位置合わせの実現が挙げられる。現在の手法では、カメラ画像の画角が過去写真と異なる場合、十分な特徴点マッチング結果が得られず、正確な位置合わせが困難である。これらの課題に対しては、SLAM を用いたトラッキングに基づく AR 提示や、過去写真から深度情報の推定を行う必要がある。これにより、過去写真の平面的な AR 提示にとどまらず、視点移動に伴う奥行きや視差を考慮した AR 提示が可能になると考えられる。

参考文献

- [1] McGookin, D.: Reveal: Investigating Proactive Location-Based Reminiscing with Personal Digital Photo Repositories, *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–14 (2019).
- [2] Bruna Costacurta Nascimento, L. C. d. S. M.: Augmented Reality and Historical Sites: A Systematic Literature Review, *Blucher Design Proceedings*, Vol. 9, No. 6, pp. 685–698 (2021).
- [3] Lee, G. A., Dünser, A., Kim, S. and Billingham, M.: CityViewAR: A Mobile Outdoor AR Application for City Visualization, *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities*, pp. 57–64 (2012).
- [4] Sauter, L., Rossetto, L. and Schuldt, H.: Exploring

- Cultural Heritage in Augmented Reality with GoFind!, *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality*, pp. 187–188 (2018).
- [5] Cavallo, M., Rhodes, G. A. and Forbes, A. G.: Riverwalk: Incorporating Historical Photographs in Public Outdoor Augmented Reality Experiences, *2016 IEEE International Symposium on Mixed and Augmented Reality*, pp. 160–165 (2016).
- [6] Blanco-Pons, S., Carrión-Ruiz, B., Duong, M., Chartrand, J., Fai, S. and Lerma, J. L.: Augmented Reality Markerless Multi-Image Outdoor Tracking System for the Historical Buildings on Parliament Hill, *Sustainability*, Vol. 11, No. 16, p. 4268 (2019).
- [7] DeTone, D., Malisiewicz, T. and Rabinovich, A.: SuperPoint: Self-Supervised Interest Point Detection and Description, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 224–236 (2018).
- [8] Sarlin, P.-E., DeTone, D., Malisiewicz, T. and Rabinovich, A.: SuperGlue: Learning Feature Matching With Graph Neural Networks, *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4937–4946 (2020).
- [9] Alcantarilla, P. F., Nuevo, J. and Bartoli, A.: Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces, *British Machine Vision Conference* (2013).
- [10] IJsselsteijn, W. A., de Kort, Y. A. and Poels, K.: The Game Experience Questionnaire, *Technische Universiteit Eindhoven* (2013).