MR2022-3

高速インスタンスセグメンテーションを用いた投影による選択的色操作の提案

小石原 遼*1　天野 敏之*2　渡辺 義浩*1

Abstract – 近年，投影型の拡張現実において，運動物体に対する色操作手法が提案されている．しかし，同手法で要請される処理時間が短いため，物体認識処理を加えることが困難であった．これに対し，シチュエーションを限定することでニューラルネットワークのモデルを軽量化し，インスタンスセグメンテーションを高速化する手法が開発された．本稿では，同手法を組み込むことで特定領域の選択的色操作を提案する．実験では，動的環境下において特定領域を抽出し，その領域に応じた補正画像を投影することで，選択的色操作が可能であることを実証した．

Keywords ： インスタンスセグメンテーション，高速ビジョン，投影型拡張現実，プロジェクタ・カメラ，高速プロジェクタ

1　はじめに

近年，プロジェクタ・カメラによって実世界を拡張する様々な研究が提案されている [1]．特に本稿では，動的に変化する実世界と仮想的な視覚情報を，人間の知覚レベルで完全に融合させるダイナミックプロジェクションマッピングに着目する．同事例では，映像の投影位置と対象の位置が一致せず，投影ずれが知覚されることが問題となっている．このようなずれを知覚されないためには，画像取得から映像投影までの処理を約 6ms 以内に収める必要がある [2]．

ダイナミックプロジェクションマッピングには，レイトレーシングを用いて高い現実感を再現する手法 [3] や，顔追跡を用いたバーチャルメイクアップ [4] などがある．本稿では，その中でも投影によって運動する現実物体に対して色操作を行う手法に着目する．

同技術では，補正画像を生成する際に対象物体の白色照明下画像が必要となる．同画像を得るために，天野らは色操作されたシーンを撮像した画像から，白色照明下画像を推定する手法を提案した [5]．この他にも，高速化に向けて，約 1000fps で投影可能な高速プロジェクタと高速カメラから成る動的環境下のための色操作手法も提案されている [6]．さらに同事例では，高速プロジェクタによる瞬間的な白色投影を補正画像の投影パターンに埋め込んで同期撮像を行うことで，推定を用いずに直接白色照明下画像を取得している．しかし，これらの手法は投影範囲全体を一様に色操作するため，対象ごとの選択的な操作が難しかった．

そこで，動的環境下における対象物体の選択的色操作を行う手法を考える．このような色操作によって，対象ごとに色操作の内容を変えることができる．本手法は，対象領域のみの色操作やラベル投影による視線誘導や UI 応用，靴やバッグなど特定のアイテムのみの色操作によるファッション分野への応用が期待できる．

選択的色操作を実現するためには，入力画像に対しピクセル単位で物体のカテゴリを識別するインスタンスセグメンテーションが有効である．しかし，既存のインスタンスセグメンテーションは速度が最大で 45fps であり [7]，これまでに述べた動的環境下における投影遅延の要請を満たすことが困難である．これに対し，我々はシチュエーションを限定した専門モデルを作成することで軽量化し，インスタンスセグメンテーションを高速化する手法を提案した [8]．

そこで本稿では，高速インスタンスセグメンテーション手法を用いて，投影による選択的色操作を提案する．具体的には，対象物体の特定領域をインスタンスセグメンテーションを用いて抽出し，その領域に応じた補正画像を投影することで選択的色操作を行う．また，インスタンスセグメンテーションを Python でなく C++で実装することで，より高速化できることを確認する．さらに，本手法を 2 つの実験で検証し，選択的色操作の有効性を示す．

2　関連研究

2.1　インスタンスセグメンテーション

He らは，Convolutional Neural Network (CNN) を用いた物体検出手法にセグメンテーション用のブランチを加えることで，インスタンスセグメンテーションを実現した [9]．しかし，本手法は候補領域を抽出するネットワークと，分類・セグメンテーションを行うネットワークが分離した Two-stage 型であり，処理量

*1 東京工業大学
*2 和歌山大学

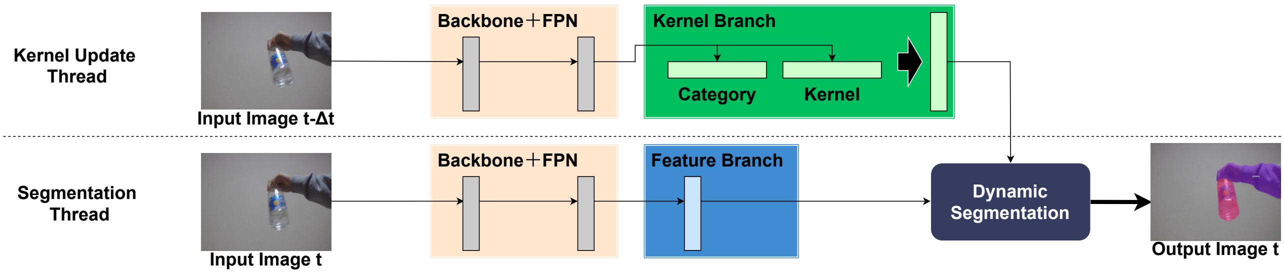

図 1　高速インスタンスセグメンテーション手法 [8]
Fig. 1　Fast instance segmentation [8]

が大きいため，速度が約 6fps と低速であった．

そこで，Daniel らはインスタンスセグメンテーションを Prototype と呼ばれる，インスタンスに依存しない特徴量マップの線形和で実現することで，候補領域抽出のステージを排除した Single-stage 型を実現した [10]．これにより，30fps の実行速度を実現した．しかし，従来の手法に比べて精度が低い欠点がある．

これに対し，Wang らは前述の手法と同程度の速度を保ちつつ，精度を向上させる手法として，SOLOv2 を提案した [7]．この手法では，バックボーンネットワークから得られた特徴量マップを 2 つのブランチに分ける．1 つ目のカーネルブランチでは，得られた特徴量マップの各層をグリッド分割し，グリッドごとに物体のカテゴリとマスク抽出のためのカーネルを出力する．もう一方の特徴量ブランチでは，得られた特徴量マップをマスク特徴量に変換する．このようにインスタンスセグメンテーションの最終的なマスク出力を，グリッドベースのカーネルを用いて検出することで，45fps 程度の速度を達成しつつ，精度を向上させた．しかし，投影型拡張現実に適用するには，さらなる高速化が必要である．

そこで我々は，選択的色操作を始めとした応用展開の多くでは，認識を必要とする対象のカテゴリ数が少なく，対象同士が密集していないこと利用して，さらなる高速化を行った [8]．具体的には，SOLOv2 をもとに図 1 のような，限定的なシチュエーションのみに対応する専門モデルを作成することでインスタンスセグメンテーションの処理量を削減した．我々は同手法を用いて，約 300fps でインスタンスセグメンテーションを実行できることを示し，投影型拡張現実への応用可能性を示した．

2.2　プロジェクタ・カメラを用いた投影型色操作に関する研究

Akiyama らは，動的に変化する環境光下で白色照明下画像を推定することで色を自由に操作できる手法を提案した [11]．しかし，高速化を目的とした手法ではなく，使用した推定手法やプロジェクタが低速だったため，対象が運動したときに，対象と補正画像の間で投影ずれが生じる問題があった．

これに対し，柳澤らは，高速プロジェクタに白色画像を埋め込み，適切なタイミングで撮像することで，推定することなく白色照明下画像を取得する手法を提案した [6]．これにより彼らは，高速なプロジェクタ・カメラを用いて，動的に変化する環境下で対象物体の色を操作できる手法を実現した．また彼らは，赤外域での高速な深度計測を組み込んだシステムへの拡張と，深度情報を用いた色操作手法も提案した [12]．しかし，これらの手法は投影範囲全体を一様に色操作するため，対象ごとの選択的な操作が難しかった．

3　高速インスタンスセグメンテーションを用いた投影による選択的色操作

3.1　概要

本節では，図 1 で示した高速インスタンスセグメンテーション手法を用いた投影による選択的色操作手法を提案する．図 2 に本手法の概要を示す．

まず，色操作を行う対象に対し，プロジェクタ・カメラを用いて白色照明下の入力画像 I を取得する．撮像は可視光プロジェクタ・カメラ間で同期をとり，白色投影のタイミングで行うことで同画像を得ることができる [6]．その後，入力画像 I に対し，高速インスタンスセグメンテーション手法 [8] を適用する．本手法を適用することで，検出した N 個のインスタンス領域を示す N 枚の抽出画像 $S=(S_1,S_2,...,S_N)$ と，各インスタンスのカテゴリ $C=(c_1,c_2,...,c_N)$ を得ることができる．次に各抽出画像 S_n の画素値を用いて，補正画像 P_n を生成する．この処理はインスタンス毎のカテゴリに応じて行われる．次に，生成された N 個の補正画像を合成し投影画像 P を作成する．最後に同画像をプロジェクタへ転送し，対象に投影することで選択的色操作を完了する．以下に，各ステップの詳細を述べる．

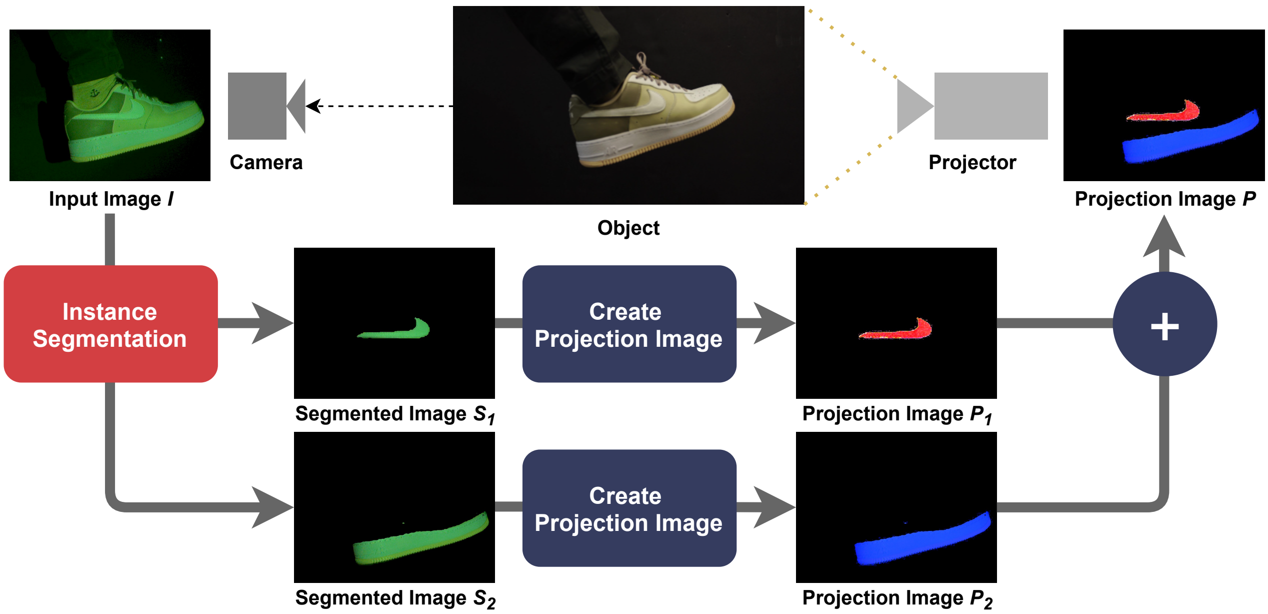

図 2　投影による選択的色操作の概要．図は検出インスタンス数が 2 の場合を示す．
Fig. 2　Overview of selective color manipulation．Figure shows the two instances example.

3.2　インスタンスセグメンテーションを用いた抽出画像の取得

本ステップでは，得られた白色照明下の入力画像 I に対して，図 1 で示した高速インスタンスセグメンテーション手法を適用して，投影したい領域のみを抽出した画像を得る．本ステップは入力画像 I からインスタンス領域を抽出した抽出画像 S と，各インスタンスのカテゴリ名を保持したカテゴリ C を出力する．このとき，検出したインスタンスが N 個の場合，抽出画像は $S = (S_1, S_2, ..., S_N)$，カテゴリは $C = (c_1, c_2, ..., c_N)$ と表せる．ここで，抽出画像 S_n において，インスタンス領域は入力画像 I の画素値，その他の領域は黒の画素値を持つ．また，抽出画像上にインスタンス n のカテゴリ名を重畳描画したものを S_n として利用することもできる．

3.3　補正画像と投影画像の生成

本手法は検出したインスタンス毎に色操作内容を選択することができる．そのため補正画像は，インスタンス毎に生成する．インスタンス n に対する補正画像 P_n は，抽出画像 S_n を作成した後，柳澤らの手法 [12] をもとに以下のように生成する．

$$K = \frac{I}{W} \tag{1}$$

$$P_n = V^{-1}(K^{(-1)} f_n(S_n) - E) \tag{2}$$

W は白色照明下画像の撮像のための瞬間的な白色投影，K は投影対象の反射率，E は環境光，V はカメラとプロジェクタ間応答の線形変換，$f_n(S_n)$ はインスタンス n の色操作の目標画像を示す．ここで生成した補正画像を投影することで，インスタンス n の領域に対して f_n で決定した色操作を行うことができる．また，実験では V, E はあらかじめキャリブレーションを行なって取得し，インスタンス n の領域外に対する環境光は 0 として計算した．

次に，複数の領域を同時に色操作を行うために，各補正画像を以下のように合成し，投影画像 P を生成する．

$$P = \sum_{n=1}^{N} P_n \tag{3}$$

同処理は，各インスタンス領域が重複していないことを前提としている．最後に同画像をプロジェクタへ転送し，投影することで対象ごとの選択的色操作を完了する．

4　実験

前節で示した選択的色操作手法を 2 つの実験で検証し，本手法の有効性を示す．実験 1 では新聞画像に含む人間領域に対する彩度減衰を行う．本実験は 1 つの対象のみ色操作やラベル投影を行うことで，観察者の視線誘導応用や UI 応用を想定して実験する．実験 2 では靴のロゴやソール領域に対する色操作を行う．本実験は，各パーツの色を変えることで，靴の履き替えの不要な試着体験の提供などのファッション応用を想定して実験する．

4.1　インスタンスセグメンテーションモデルの作成

本手法では，抽出画像の生成に我々が開発したインスタンスセグメンテーション手法 [8] を用いて実装す

表 1　処理時間 (ms)
Table 1　Processing time (ms)

	Backbone+FPN	Feature branch	Dynamic segmentation	合計
Python[8]	1.474	0.492	1.756	3.722
C++	0.468	0.196	0.586	**1.250**

図 3　実験 1: 左から順に，(a) 入力画像 I，(b) 抽出画像 S_1，(c) ラベル付き抽出画像
Fig. 3　Experiment 1: From the left, (a) Input image I, (b) Segmented image S_1, (d) Segmented image with label

図 4　実験 1: 左から順に，(a) 色操作なし，(b) インスタンス領域の彩度減衰（P_1 投影），(c) ラベル投影
Fig. 4　Experiment 1: From the left, (a) Uncorrected object, (b) Decreasing color saturation (Project P_1), (c) Label projection

る．同手法は各実験に対応するモデルを事前に用意する必要がある．同モデルを作成するにあたって，白色照明下の対象物体の画像を撮像し，各画像の各対象領域に対するアノテーションを用いてデータセットを作成する．実験 1 で学習するデータセットは，白色照明下の対象物体を，500fps で撮像した画像 1000 枚を用い，[8] と同様に作成した．検出するインスタンスは新聞紙上の人間とする．実験 2 で学習するデータセットは，白色照明下画像 100 枚をインタラクティブセグメンテーション [13] を用いて作成した．検出するインスタンスは靴のロゴとソールとし，合計 3 つのモデルを作成する．ここで，ロゴのみを検出するモデル 1，ソールのみを検出するモデル 2，ロゴ・ソールを同時に検出するモデル 3 とする．また，学習は Python，推論には C++で実装して実験を行った．

4.2　実験環境・条件

本システムは，最大 947fps でのカラー投影が可能な高速プロジェクタ [14](解像度 1024×768) と最大 520fps でカラー撮像が可能な高速カメラ (Basler acA720-520 uc，解像度 720×540) を用いて構築した．また，演算，及び画像処理を行う計算機は，CPU に Xeon Gold 6234 3.3GHz dual を，GPU に GeForce RTX 3090 を搭載したものを用いた．実験において，補正画像に埋め込んだ瞬間白色投影の投影時間は 200us とした．

4.3　実験結果

インスタンスセグメンテーションの処理時間を表 1 に示す．推論の実装に Python を用いた手法 [8] と比較して，処理時間を約 2.5ms 短縮できることを確認した．

実験 1 において，インスタンスセグメンテーションを用いて取得した抽出画像を図 3 に示す．左から順にそれぞれ，(a) は入力画像 I，(b) はインスタンス領域の抽出画像 S_1，(c) はインスタンス領域のラベル付き抽出画像である．(b) の結果から，(a) のような入力に対し，人間の領域が適切に取得できていることが確認できる．また，(c) の結果から，インスタンスのカテゴリを適切に検出できていることが確認できる．

また，図 3 で抽出した画像を用いて生成した補正画像を，色操作領域に投影した結果を図 4 に示す．左から順にそれぞれ，(a) は投影なし，(b) は抽出画像 S_1 を用いて生成した補正画像 P_1 を投影した結果，(c) はラベル付き抽出画像を用いて生成した補正画像の投影結果である．(b) では投影結果が彩度減衰画像，(c) ではインスタンス領域をオレンジ色，ラベルを緑色となるように補正画像を生成した．本実験では，検出したインスタンスは 1 つだけであるので，補正画像の合成は行っていない．またこれらの結果から，対象領域を色操作できることを確認した．

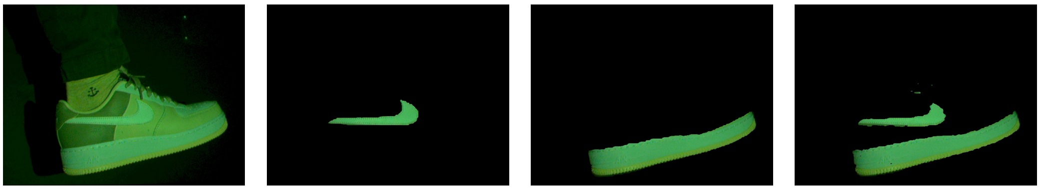

図 5　実験 2: 左から順に，(a) 入力画像 I，(b) モデル 1 のロゴ抽出画像 S_1，(c) モデル 2 のソール抽出画像 S_1，(d) モデル 3 のロゴ・ソール抽出画像 $S_1 + S_2$

Fig. 5　Experiment 2: From the left, (a) Input image I , (b) Segmented logo image S_1 of model1, (c) Segmented sole image S_1 of model2, (d) Segmented logo and sole image $S_1 + S_2$ of model3

図 6　実験 2: 左から順に，(a) 色操作なし，(b) ロゴ色操作，(c) ソール色操作，(d) ロゴ・ソール色操作

Fig. 6　Experiment 2: From the left, (a) Uncorrected object, (b) Logo's color manipulation, (c) Sole's color manipulation, (d) Logo and sole's color manipulation

次に，実験 2 において，インスタンスセグメンテーションを用いて取得した抽出画像を図 5 に示す．左から順にそれぞれ，(a) は入力画像 I，(b) はモデル 1 の出力であるロゴ領域の抽出画像 S_1，(c) はモデル 2 の出力であるソール領域の抽出画像 S_1，(d) はモデル 3 の出力であるロゴ領域の抽出画像 S_1 とソール領域の抽出画像 S_2 を一つの画像にまとめたものである．(b), (c) の結果より，一つのインスタンスのみを検出する際は，適切に領域を抽出できていることが確認できた．しかし，(d) の結果より，2 つのインスタンスを検出する際は少し精度が低くなることを確認した．

最後に，図 5 で抽出した画像を用いて生成した補正画像を，色操作領域に投影した結果を図 6 に示す．左から順にそれぞれ，(a) は投影なし，(b) はモデル 1 の抽出画像 S_1 を用いて生成した補正画像 P_1 を投影した結果，(b) はモデル 2 の抽出画像 S_1 を用いて生成した補正画像 P_1 を投影した結果，(d) はモデル 3 の抽出画像 S_1，S_2 を用いて生成した投影画像 $P = P_1 + P_2$ を投影した結果である．ロゴの補正画像は投影結果が赤色，ソールの補正画像は投影結果が青色になるように生成した．これらの結果から，複数領域を選択的に色操作できることを確認した．また，動的な物体に対して，遅延の少ない色操作ができることを確認した．

5　考察

本実験で，対象領域の色を選択的に色操作できることを確認できた．また，本稿では 2 つの色操作応用を想定して実験を行ったが，本手法は他にも様々な応用例が考えられる．そのうちの 1 つとして，花や料理の色操作応用が挙げられる．本手法を用いて，特定の料理や特定の花の領域を抽出し，補正画像を投影することで，これらの見かけを操作できる．また，作業支援への応用も考えられる．作業対象の領域を抽出し，その領域に作業工程の動きを投影することで，支援を実世界で行うことができる．

一方，図 6 の (d) に示されるように，色操作においてセグメンテーションの精度が不足していることを確認した．従って，本手法を様々なシチュエーションに応用する際に，インスタンスセグメンテーションの精度向上が不可欠である．精度不足の原因として，検出対象のインスタンスが増えた際に，学習のための画像枚数が足りていないためと考えられる．しかし，モデルを作成するたびに学習画像を大量に収集する場合，学習準備にかなりの時間を要してしまう．今後は，少ない画像枚数でも多くの情報量を持たせることができるようなデータセット拡張を適用することで，精度のさらなる向上が望めると考える．

6　まとめ

本論文では，高速なインスタンスセグメンテーション手法を用いた投影による選択的色操作の提案を行った．この実現にあたり提案した手法では，取得した白色照明下画像に対し，高速なインスタンスセグメンテーションを用いて対象領域を抽出した．この各対象領域に対し，色操作を行うための画像である投影画像を生成し，その画像を投影することで対象領域の選択的な色操作を行った．また，実験により動的な各対象領域を選択的に色操作ができることを確認した．

参考文献

[1] A. Grundhöfer, and D. Iwai, "Recent advances in projection mapping algorithms, hardware and applications," *Computer Graphics Forum*, 2018.

[2] A. Ng, J. Lepinski, D. Wigdor, S. Sanders, and P. Dietz, "Designing for low-latency direct-touch input," in *UIST*, 2012.

[3] T. Nomoto, R. Koishihara, and Y. Watanabe, "Realistic dynamic projection mapping using real-time ray tracing," in *SIGGRAPH*, 2020.

[4] A. H. Bermano, M. Billeter, D. Iwai, and A. Grundhöfer, "Makeup lamps: Live augmentation of human faces via projection," *Computer Graphics Forum*, 2017.

[5] 天野敏之，加藤博一，"モデル予測制御を用いたプロジェクタカメラ系による見かけの制御"，電子情報通信学会論文誌，2011.

[6] 柳澤昂輝，野元貴史，渡辺義浩，"高速プロジェクタ・カメラを用いた動的環境下における見かけ操作の基礎検証"，in *SIG-MR*，2020.

[7] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, "Solov2: Dynamic and fast instance segmentation," in *NIPS*, 2020.

[8] 小石原遼，渡辺義浩，"投影型拡張現実への応用に向けたインスタンスセグメンテーションの高速化"，in *MIRU*，2021.

[9] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *ICCV*, 2017.

[10] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "Yolact: Real-time instance segmentation," in *ICCV*, 2019.

[11] R. Akiyama, G. Yamamoto, T. Amano, T. Taketomi, C. Sandor, and H. Kato, "Appearance control in dynamic light environments with a projector-camera system," in *IEEE VR*, 2016.

[12] 柳澤昂輝，野元貴史，角野究，上田晋寛，加瀬部秀訓，山下徹，湯浅剛，天野敏之，渡辺義浩，"高速プロジェクタ・カメラとリアルタイム 3 次元計測を用いた実世界の色操作"，in *SIG-MR*，2021.

[13] K. Sofiiuk, I. A. Petrov, and A. Konushin, "Reviving iterative training with mask guidance for interactive segmentation," 2021.

[14] Y. Watanabe, and M. Ishikawa, "High-speed and high-brightness color single-chip DLP projector using high-power LED-based light sources," in *IDW*, 2019.