

Augmented Realityの動的オクルージョン解決に向けた リアルタイムセマンティックセグメンテーションシステムの開発

○城戸 大輝^{*1} 福田 知弘^{*2}
矢吹 信喜^{*3}

キーワード：拡張現実（Augmented Reality） オクルージョン セマンティックセグメンテーション
深層学習 景観検討

1. 研究の背景と目的

魅力的な都市景観の形成には、事業に携わる関係者間で合意形成を図り、相互連携して事業を進めることが必要である。利害関係者には、行政機関や専門家だけでなく、住民や一般利用者も含まれるため、様々な専門性や知識レベルの差が予想される。そのため、完成後に予想される景観を可視化し、利害関係者全体で事業内容を適切に共有することが重要である。

景観検討手法として、拡張現実（Augmented Reality：AR）を用いた手法が着目されている^{1),2)}。ARは現実空間に情報を重畳する技術であり、視覚的なARはカメラからリアルタイムに取得された映像に3-Dimensional（3D）モデルを重畳することで実現する。ARによる景観検討では、新設構造物の周辺環境を実写映像で表現するため、仮想現実（Virtual Reality：VR）の課題である周辺環境のモデリングの手間やコストを削減できる。また、ARによる景観検討は、実空間との繋がりを維持することが可能である。事業予定地で新設構造物の3Dモデルを実際のスケールで重畳させて確認するため、会議室での検討と比較して、より臨場感のある景観検討を可能にする。ARの実現において生じる問題として、現実空間と仮想物体の前後関係を正しい順番に表現できないオクルージョン問題がある³⁾。ARはカメラから取得した実画像上に3Dモデルを重畳し表示することで実現するため、手前側に存在するはずの実物体が合成された3Dモデルに隠れることで生じる。オクルージョンは奥行き情報を示す重要な指標であり、正確に表現することが求められる。

本研究では、このオクルージョン問題について着目する。既往研究⁴⁾として、Inoueらは、ARと隠消現実（Diminished Reality：DR）による景観検討を実現するPhotoAR+DR2017を開発した。DRは実世界に存在する物体に対し、裏側の背景（以降、隠背景とする）を重畳することで視覚的消去を実現する技術⁵⁾である。PhotoAR+DR2017では、複数の写真群から3Dモデルを復元する技術であるStructure from Motion（SfM）を用いて隠背景モデルを作成し、既設構造物を視覚的に消去することで、既設構造物の解体撤去前でもARによる適切な

景観検討を可能にした。このシステムでは、ARにおけるオクルージョン問題に対し、SfMを用いて事前にオクルージョン表現を行う実物体のモデル（以降、オクルージョンモデルとする）を作成し、ARモデルに隠れる部分についてARモデルをレンダリングしないことで、オクルージョン表現を実装している。しかしこの手法では、木々のように実物体が時間によって変化する場合、オクルージョンモデル作成時とシステム使用時のタイムラグにより、モデルと実物体の形状に差が生じ、適切なオクルージョン表現が難しい。また、ARによるシミュレーション中に歩行者や車が通った場合、事前にオクルージョンモデルを作成することは困難であるため、オクルージョン表現は不可能である。したがって、リアルタイムで物体を認識し、オクルージョン処理を行うことが求められる。また、筆者らの既往研究⁶⁾として、移動物体を含むDRシステムであるDRwOD2018を開発した。このシステムでは、深層学習ベースの物体検出技術によってリアルタイムで人や車等の移動物体を検出し、視覚的消去を実現した。このシステムを応用することで動的オクルージョン表現は可能であると考えられるが、物体検出の際、物体を含む矩形での検出しか実施できないという制約が残っており、正確なオクルージョン表現は実現できていない。

本研究の目的は、動的なオクルージョン表現の実装に向けた、リアルタイムセマンティックセグメンテーションシステムを構築することである。セマンティックセグメンテーションは、画像に対し画素単位で識別、ラベル付けを行う技術である。これにより、特定の物体の領域のみの抽出が可能となり、動的オクルージョン表現への応用が可能と考えられる。本研究では、リアルタイム性とセグメンテーションの精度を考慮し、セマンティックセグメンテーション手法として深層学習ベースのICNet⁷⁾を使用する。ICNetは高解像度の入力画像に対し、リアルタイムで高精度なセグメンテーションを行うことを目的に開発された手法である。ICNetによって入力画像のセグメンテーションを行い、オクルージョン表現を行う物体のマスク画像を作成することで、適切なオクルージョン表現を実現する。検証実験として、構築したシステムの動作確認を実施した。

2. セマンティックセグメンテーションシステムの構築

セマンティックセグメンテーションシステムの構築において、リアルタイムでのセグメンテーション処理は非常に負荷の大きい処理であり、高性能なデスクトップ PC を要する。しかし、本システムは屋外での景観検討への適用を想定しているため、ノート PC やタブレットといったモバイル端末での使用が予想される。そこで本研究では、クライアント側で取得した動画をサーバに転送し、サーバ側でセグメンテーション処理を施した画像を再びクライアント側に転送するシステムを構築した。図 1 にシステム利用図を示す。

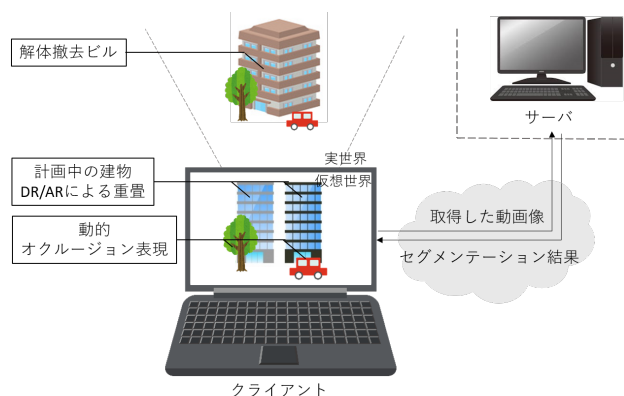


図 1 システム利用図

2.1. サーバ・クライアント間の通信

サーバ・クライアント間の通信において、本研究ではプログラミング言語 Python 用の軽量な Web アプリケーションフレームワークである Flask を使用した。システムの開発環境は、PhotoAR+DR2017⁴⁾が Graphical User Interface (GUI) ベースでの AR・DR システムの構築が可能であるゲームエンジンの Unity 上で開発されていることを継承し、Unity 上でシステムを構築した。

初めに、サーバ側の PC で Flask を用いた Web アプリケーションの開発を行った。この Web アプリケーションでは、指定したホストのポートにアクセスしたクライアント側の PC から画像データを取得し、ICNet によるセグメンテーション処理を施し、再びクライアント側に転送する処理を実装した。次に、クライアント側の PC の Unity で、Web カメラから取得したフレームを Web アプリケーションを起動しているサーバ側の PC に転送し、セグメンテーション処理が施された画像を取得し表示する機能を実装した。この処理の実装には、Unity における URL からコンテンツを取得するモジュールの www クラスを使用した。図 2 にサーバ・クライアント間のデータ通信のフローを示す。

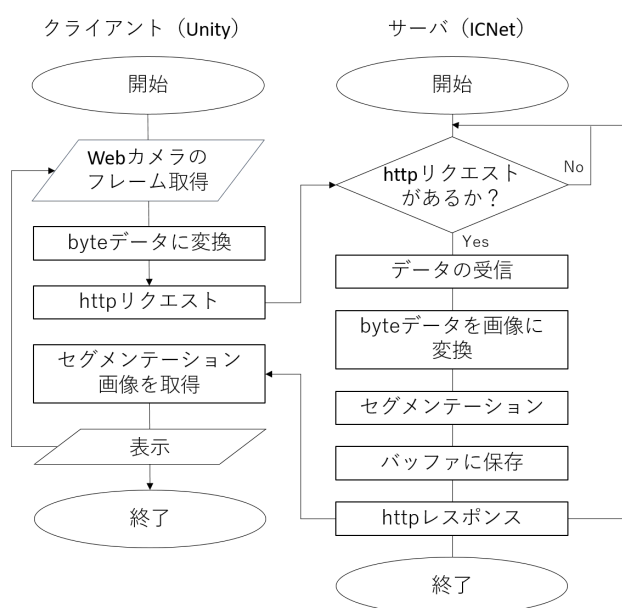


図 2 サーバ・クライアント間の通信フロー

2.2. 通信のタイムラグによる表示のギャップの解消

クライアント側の PC の Unity で Web カメラのフレームを取得してからセグメンテーション画像を取得するまで、サーバ・クライアント間の通信によるタイムラグが生じる。そのため、Web カメラから取得した動画をそのまま表示すると、セグメンテーション画像の表示とズレが生じ、正しいオクルージョン表現が困難になることが予想される。そこで、この取得した動画とセグメンテーション画像の表示のギャップの解消を実施した。

現状、この通信のタイムラグ自体を解消することは不可能である。そこで本システムでは、Unity 上での取得した動画の表示について、サーバに転送した画像とセグメンテーション画像の表示を対応させることで疑似的に表示のギャップを解消した（図 3）。図 3 から、(a) と (c) は入力画像とセグメンテーション画像が対応していないが、(b) と (c) は対応していることが確認できる。

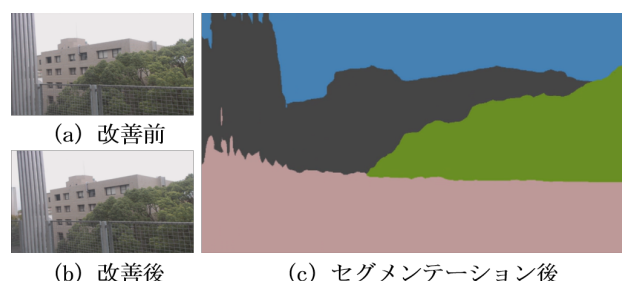


図 3 Unity 上における表示のズレの解消

2.3. 動的オクルージョンレイヤーの実装

AR における動的オクルージョン表現の実装にあたり、既存のシステムに対し、新たに動的オクルージョン表現を

実装する動的オクルージョンレイヤーを構築した。図 4 に動的オクルージョンに対応した AR による景観検討を実現するレイヤー構造を示す。PhotoAR+DR2017 は、事前にオクルージョンモデルを作成し、AR モデルとオクルージョンモデルの位置から AR レイヤーにおける AR モデルのレンダリングを調整することで、オクルージョン表現を実装した。本システムでは、動的オクルージョンレイヤーを実装し、動的オクルージョン表現を行う実物体を投影することで、AR による動的オクルージョン表現を実装する。

動的オクルージョンレイヤーの実装にあたり、まず動的オクルージョン表現を行う実物体のマスク画像を作成する。この作成したマスク画像に対して現況動画を合成することで、動的オクルージョン表現を行う物体のみを投影することが可能となる。マスク画像作成において、Unity 用プラグインである OpenCV for Unity を使用した。セグメンテーション画像に対し、既に定義されている特定のラベルの領域を白 ((R,G,B) = (255, 255, 255))、それ以外のラベルの領域を黒 ((R,G,B) = (0, 0, 0)) で描画することにより、特定のラベルに対するマスク画像作成を実装した (図 5 (c))。また、このマスク画像に対して現況動画を合成することで、特定のラベルの物体のみの投影を実装した (図 5 (d))。

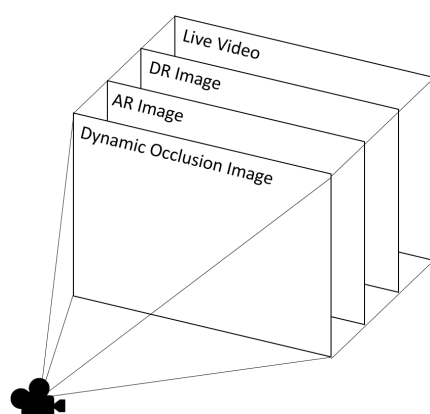


図 4 動的オクルージョンに対応したレイヤー構造

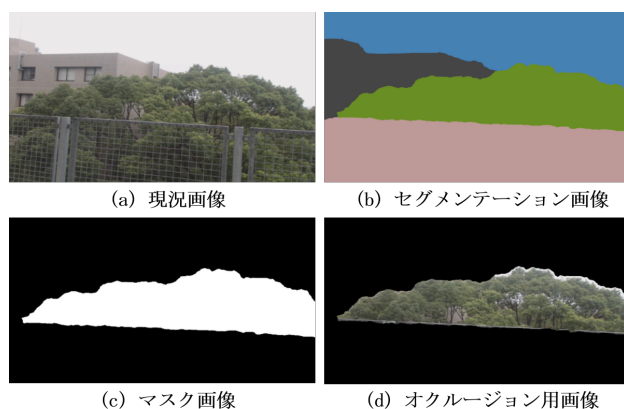


図 5 マスク画像の作成及び現況画像の合成

3. 検証実験

開発したシステムの動作確認を実施した。サーバ側のデスクトップ PC は、CPU Intel Core i7-8700K, GPU NVIDIA GeForce GTX 1080Ti 11GB, RAM 32G を使用し、クライアント側のノート PC は、CPU Intel Core i7-7700HQ, GPU NVIDIA GeForce GTX1060 6GB, RAM 8G を使用した。ICNet によるセグメンテーションの学習モデルには、Cityscapes データセット⁸⁾を使用した。Cityscapes はドイツの 50 の都市の画像にセマンティック・セグメンテーションの情報が付加された大規模データセットであり、5,000 枚の細かいピクセルレベルでセグメンテーション情報が付加された画像と 20,000 枚の粗いピクセルレベルの情報が付加された画像で構成されている。Cityscapes を採用した理由は、都市のシーン理解のために制作されたデータセットであり、本研究の目的にも沿っているためである。本システムにおける検出対象のラベル一覧を図 6 に示す。また、現状ではサーバ側のデスクトップ PC とクライアント側のノート PC の通信は LAN (Local Area Network) 内でのみ可能であるため、LAN 内で通信を実施した。ノート PC に接続した Web カメラを横に回転させ、現況動画を取得した。なお、本検証では、検出対象を vegetation とし、現況動画、セグメンテーション画像、マスク画像、マスク画像と現況画像を合成したオクルージョン用画像を表示した。現況動画の解像度は 1024×576 [pixels]とした。また、動作速度を確認するため、フレームレートを表示した。検証結果を図 7 に示す。

検証の結果、現況の動画から vegetation の検出及びマスク画像の作成を行い、現況画像と合成することで、動的オクルージョンのための vegetation のみの表示が可能であることを確認した。システムの動作速度は約 5fps (frame per second) であり、リアルタイム描画を考慮すると、不十分な値となった。クライアント側で取得した現況動画を表示する処理は約 30fps、サーバ側の ICNet によるセグメンテーション処理は約 70fps ほどであったため、システムの動作速度が不十分であった原因は、サーバ・クライアント間の通信速度に問題があると考えられる。現在は LAN 内で通信しているが、WAN (World Area Network) / インターネット環境への適用を考慮すると、現状以上に動作速度への影響が予想される。したがって、通信速度は改善すべき点であることが確認できた。

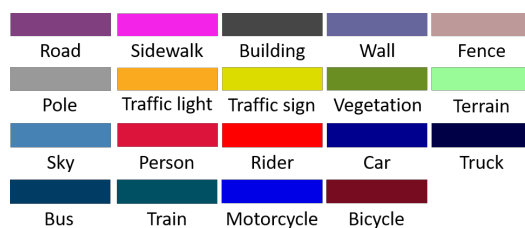


図 6 検出対象のラベル一覧

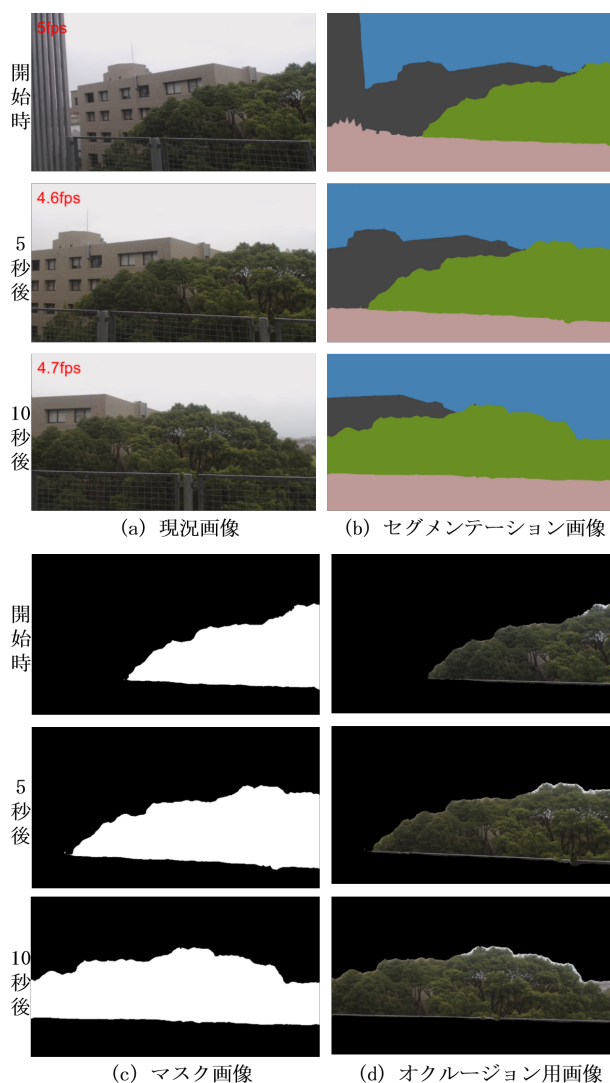


図 7 検証結果

4. 結論と今後の展望

本研究では、AR におけるオクルージョン問題について、動的なオクルージョン表現を実装するためのリアルタイムセマンティックセグメンテーションシステムの構築を行った。本研究の結論を以下に述べる。

- ・ クライアント側の Web カメラで取得した動画をサーバに転送し、サーバ側でセマンティックセグメンテーション処理を施した画像をクライアント側に転送するシステムを構築した。
- ・ 画像処理技術により、現況画像とセグメンテーション後の画像から特定のラベルの物体のみを表示する機能を実装した。
- ・ 検証実験を実施し、特定の物体の動的なオクルージョンレンダリングが可能であることを確認した。

今後の展望として、本システムと Inoue らが開発した PhotoAR+DR2017 を統合することで、AR による景観検討における動的オクルージョン表現を実装する。さらに、WAN/インターネット環境への適用も検討していく。また、

筆者らが開発した DRwOD2018 では矩形検出しか実現できない課題に対しても、本研究で開発したシステムを適用することで、対象領域のみの抽出が実現できると考えられる。

謝辞

本研究の一部は、JSPS 科研費 JP16K00707 の助成を受けたものである。謹んで感謝の意を表する。

[参考文献]

- 1) Klinker, G., Stricker, D. and Reiners, D.: 2001, Augmented Reality for Exterior Construction Applications, in W. Barfield and T. Caudell (eds.), Augmented Reality and Wearable Computers, Lawrence Erlbaum Press, New Jersey, pp.397-427.
- 2) Yabuki, N., Miyashita, K., and Fukuda, T.: 2011, An Invisible Height Evaluation System for Building Height Regulation to Preserve Good Landscapes using Augmented Reality, Automation in Construction 20 (3), pp.228-235.
- 3) Shah, MM., Arshad, H., and Sulaiman, R.: 2012, Occlusion in augmented reality, Proceedings of the 8th International Conference on Information Science and Digital Content Technology (ICIDT 2012), 372-378.
- 4) Inoue, K., Fukuda, T., Cao, R., and Yabuki, N.: 2018, Tracking Robustness and Green View Index Estimation of Augmented and Diminished Reality for Environmental Design: PhotoAR+DR2017 project, Proceedings of the 23rd International Conference on Computer-Aided Architectural Design Research in Asia (CAADRIA 2018), 339-348.
- 5) Mori, S., Ikeda, S., and Saito, H.: 2017, A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects, IPSJ Transactions on Computer Vision and Applications, 9, 1-17.
- 6) Kido, D., Fukuda, T., and Yabuki, N.: 2018, Development of Real-time Object Detection System with Diminished Reality for Landscape Design Simulation, International Conference on Asian-Pacific Planning Societies (APPS 2018), 223.
- 7) Zhao, H., Qi, X., Shen, A., Shi, J., and Jia, J.: 2017, ICNet for Real-Time Semantic Segmentation on High-Resolution Images, arXiv preprint arXiv:1704.08545.
- 8) M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele.: 2016, The Cityscapes Dataset for Semantic uBan Scene Understanding, In CVPR.

- *1 大阪大学 大学院工学研究科 環境・エネルギー工学専攻 博士前期課程
- *2 大阪大学 大学院工学研究科 環境・エネルギー工学専攻 准教授 博士(工学)
- *3 大阪大学 大学院工学研究科 環境・エネルギー工学専攻 教授 Ph.D.