深層学習を用いたARによる道路情報提示のための学習データの自動生成に関する検討

A Note on Generation of Data for Transportation Data Indication on AR by Utilizing Deep Learning

北海道大学　○学生会員　阿部恭征 (Takayuki Abe)
北海道大学　正会員　高橋翔 (Sho Takahashi)　北海道大学　正会員　萩原亨 (Toru Hagiwara)

1. はじめに

計算機の小型・高性能化により，屋外で利用可能な小型の計算機が普及している．これに伴い，拡張現実（以降, AR）による視覚的な情報提示を交通利用者に対して行うことが可能となってきている[1]. AR とは，実世界に情報を付加し，人間の五感を拡張する技術である．特に AR による情報提示は，透過型のディスプレイなどによって情報を視野に重畳することで行われる．このため，ユーザーは視線を維持したまま情報が取得可能となり，また，ヘッドマウントディスプレイ（以降, HMD）等のデバイスを用いることで，ハンズフリーでの情報の取得が可能となる．この AR を道路利用者への情報提示に応用することで，より安全かつ効果的な道路利用が期待できる．しかしながら，従来の AR は，道路利用者に情報を提示する場合，道路利用者が視認すべき重要な実物体が遮蔽される可能性がある．

そこで，以前に著者らは，AR の導入による交通の高度化に向け，Content-aware な画像処理手法である顕著性マップのモデル[2]と Seam Carving のモデル[3]を用いて情報提示可能な位置を推定する手法を提案し，その有効性を確認した[4]．しかしながら，文献[4]の手法は，入力する動画像の画素数などに応じたパラメータを設定や，顕著性マップや Seam Carving 以外の手法を導入した高精度化を図る際に，計算コストの面での課題が大きくなり，実時間での利用が困難となる．

一方，昨今のコンピュータビジョンの発展に伴い，深層学習を用いた画像認識が注目されており，様々なタスクの精度向上が報告されている．深層学習は，一般に，入力画像と対応するラベルの組を用い，多量の隠れ層を構築することで弁別性の高い特徴算出および高精度な識別を可能としている.しかしながら，充分な精度となる深層学習のモデル構築には，膨大な量の学習データが必要であり，これを自動で用意可能とすることが大きな課題の一つである．

そこで本研究では，文献[4]によって大量に学習データを自動生成し，それを学習する識別器を構築する手法を提案する．特に本稿では，文献[4]が大量の学習データを必要とする深層学習の重大な問題に対して有効であることを確認する．具体的には，まず，道路利用者を想定した一人称視点映像（以降，視点映像）の各フレームから，文献[4]によって情報提示可能な位置を推定する．その推定結果をラベルとし，さらに，Inception-v3[5]を用いて画像特徴量を取得する．最後に，得られた特徴量を入力とする識別器として 3 層のニューラルネットワークからなる Extreme Learning Machine(ELM)[6]を構築し，構築された ELM を用いることで新たな入力映像に対して情報提示可能な位置を識別する．

2. 深層学習による情報提示可能な位置の識別

本章では提案手法について説明する．提案手法では，まず，視点映像の各フレームを任意サイズのパッチに分割し，それぞれを文献[4]で求める情報提示領域としての適性度に基づいて学習データのラベルを決定する．また，識別器として，Extreme Learning Machine(ELM)[6]を用い，入力画像から求める特徴量は，深層学習のモデルの一つである Inception-v3 から求める．これによって，新たな映像を入力とした際にも，Inception-v3 による特徴量を入力とする ELM によって，情報提示に適している度合いに応じたクラスの識別を可能とする．

2.1. ARによる情報提示が可能な度合いの算出

本節では，著者らが以前に提案した，AR による道路利用者への情報提示の支援を目的とした情報提示領域の推定手法[4]について説明する．道路交通を対象とした AR では，道路利用者が本来的に視認すべき実物体を遮蔽するような情報提示は避けなければならない．そこで，画像中の視覚的に顕著な領域を求める顕著性マップのモデル[2]，および情報量が少ない領域を検出する Seam Carving のモデル[3]を用いることで，情報提示が可能な領域を推定している．これによって，AR による情報提示を支援することが可能となる．

2.1.1. 顕著性マップのモデル

本項では，顕著性マップのモデル[2]について説明する．文献[2]では，画像中の視覚的に顕著な領域の検出を可能としている．また，このモデルは単一のフレームのみではなく映像を対象としている．文献[4]では，これを情報提示可能な領域の推定に導入することで，道路利用者の視認すべき重要な実物体を含む領域の検出をしている．

2.1.2. Seam Carving のモデル

本項では，Seam Carving のモデル[3]について説明する．文献[4]では画像中の視覚的な情報量が少ない領域を検出可能としている．これを用いることで，画像内に撮像

(a)　入力画像　(b)　顕著性マップ

(c)　Seam Carving　(d)　推定結果

図 1　文献[5]による情報提示可能な領域推定

されている視覚的な重要度が高い物体以外の画素が検出可能となる．したがって，文献[4]に基づいて情報提示可能な領域を推定することで，2.1.1 とは逆の道路交通において視認すべき物体以外の領域が検出可能となる．

2.1.3. 情報提示領域の推定

本項では，前項までの 2 つのモデルを導入した AR による情報提示が可能な領域の推定について説明する．文献[4]では，図 1(a)に示す画像を入力としたとき，前項までに説明した 2 つのモデルのそれぞれによって図(b)及び(c)を得る．さらに文献[4]では，これらの結果を重ね合わせることで，図 1(d)の結果を得ており，これを最終的な推定結果としている．この時，図 1(b)では黒い領域が低い顕著性となる部分となり，情報提示が可能である度合が高いことを示す．一方，図 1(c)と(d)では，情報提示が可能である度合が高い場所を白で示している．

2.2. 深層学習による識別

本説では，深層学習のモデル構築において，膨大な量の学習データが必要であるという大きな課題に対して，大量の学習データを自動生成し，それを学習することによる識別器の構築を可能とする手法について説明する．

2.2.1 Inception-v3 を用いた特徴量抽出

本節では，画像特徴量の抽出に用いる Inception-v3 について説明する．Inception-v3 は，一般物体認識用の大規模データセットである ImageNet を用いて学習された深層学習モデルの一つである．提案手法では，この Inception-v3 を特徴抽出器として用いて，入力画像を表現する 2,048 次元の特徴量を取得する．

2.2.2　ELM による識別

本節では，本稿で用いた ELM について説明する．ELM は Single hidden layer feedforward neural networks (SLFNs)の一種であり，3 層のニューラルネットワークからなる識別器である．提案手法では，学習データとして 2,048 次元の特徴ベクトル$v_i = [v_{i,1}, v_{i,2}, \dots, v_{i,2048}]^T$及びそのラベル$z_i = [z_{i,1}, z_{i,2}, \dots, z_{i,j},]^T$を用いる．また，提案手法では，最終層の重み$\beta$を以下によって求める．具体的には，シグモイド関数$G$を用いて特徴変換を行うことで，

$$\boldsymbol{h}(\boldsymbol{v_1}) = [\boldsymbol{G}(\boldsymbol{a_1}, \boldsymbol{b_1}, \boldsymbol{v_1}), \dots, \boldsymbol{G}(\boldsymbol{a_1}, \boldsymbol{b_1}, \boldsymbol{v_1})]^{\boldsymbol{T}} \quad (1)$$

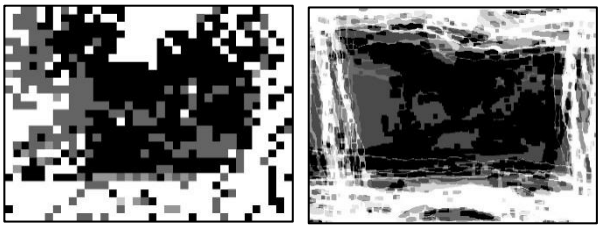
を算出する．ただし，$a_l(\mathrm{l} = 1,2,\dots,\mathrm{L})$，及び$b_l(\mathrm{l} = 1,2,\dots,\mathrm{L})$はシグモイド関数$G$のパラメータであり，L は隠れ層のノード数を表す．次に，下式によって最終層の重みβを求める．

$$\beta = (H^T H)^{-1} H^T T \quad (2)$$

ただし，$T = [z_1, z_2, \dots, z_M,]^T$，$\mathrm{H}=[h(v_1), h(v_2), \dots, h(v_M)]^T$である．最後に，テストにおいて，特徴ベクトル v を ELM に入力したとき，出力地は$f = \mathrm{h}(v)^T\beta$であり，クラスラベルは f のうち，最大値を出力したノードに対応するクラスラベルとなる．

表 1　実験動画のプロパティ

道路利用者のタイプ	縦横サイズ	フレームレート
ドライバー	720x540	30
サイクリスト	540x960	30
歩行者	540x960	30



(a)　識別結果　(b)　推定結果

図 2　提案手法による識別結果と文献[5]による推定結果

3. 実験

本章では，提案手法の有効性を確認するための実験を行う．本実験では，文献[4]で用いられた動画像を学習，およびテストデータとして用いた．動画像に関するプロパティを表 1 に示す．また，提案手法では，各フレームを 20 ピクセル四方でパッチ分割し，ランダムに 100,000 枚のパッチ状の画像を抽出した．その後，パッチ状の画像から Inception-v3 を用いて特徴量を抽出し，文献[4]の手法によって算出された情報提示領域としての適性度合いに基づいてラベルを付与することで，学習用のデータセットを生成した．このとき，ラベル数は，実験的に 5 クラスとした．

本実験では，提案手法および文献[4]の手法によって得られた結果を比較する．それぞれの手法によって得られた結果の例を図 2 に示す．図 2 を見ると，(a)の提案手法による識別結果が，(b)の文献[4]によって算出された情報提示領域としての適性度合いと類似しており，大量の学習データを自動生成し，それを学習することによって識別器を構築することの有効性が確認できる．

4. まとめ

本稿では，文献[4]によって大量に学習データを自動生

成し，それを学習する識別器を構築する手法を提案した．
また，実験によって提案手法の有効性が確認された．

謝辞：本研究の一部は，JSPS 科研費 JP17K00148, JP19H02254 および公益財団法人戸田育英財団の助成を受けて行われた．

参考文献

[1] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, “Recent advances in augmented reality,” *IEEE Comput. Graph. Appl.*, vol. 21, no. 6, pp. 34–47, 2001.

[2] B. Wang and P. Dudek, “A fast self-tuning background subtraction algorithm,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 401–404, 2014.

[3] S. Avidan and A. Shamir, “Seam carving for content-aware image resizing,” *Proc. ACM SIGGRAPH Conf. Comput. Graph.*, 2007.

[4] T. Abe, S. Takahashi, T. Hagiwara, “A Calculation Method of Degree of Data Indication Regions in First-Person View Videos for Improvement Transportation,” Proc. In the IEEE 8th Global Coferenceon Consumer Electronics, (October 2019, Accepted for presentation).

[5] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 2818–2826, 2016.

[6] G. Bin Huang, Q. Y. Zhu, and C. K. Siew, “Extreme learning machine: A new learning scheme of feedforward neural networks,” *IEEE Int. Conf. Neural Networks - Conf. Proc.*, vol. 2, no. February 2014, pp. 985–990, 2004.