

実写画像からのピクトグラム合成画像生成

綿引 優人[†] 戀津 魁[†] 柿本 正憲^{†,††}

[†]東京工科大学 メディア学部 ^{††}ドワンゴ CG リサーチ

1 はじめに

ピクトグラムは、イラストだけで情報を伝える記号言語であり、公共施設案内や発表資料など日常的によく使われている。近年、デジタルサイネージの普及に伴い、アニメーション利用を想定した拡張が期待されている。また、ピクトグラムによる視覚情報の単純化は表現手法の一つであり、芸術性があると考えられる。この芸術性に着目したデザイナーも存在し、ホームページにいくつか作品が掲載されている[1]。

ピクトグラムの需要から今までにいくつか研究されていたが、静止画像全体からピクトグラムに変換する研究は存在しない。本研究では静止画像からピクトグラム風の画像を生成することで、ピクトグラム動画制作支援の基盤になることやピクトグラム風アート制作の支援を目指す。

2 関連研究

提案手法では画像全体を変換するため、画像のどこに何があるのか認識する必要がある。そこで効率よく画像認識をするためにセマンティックセグメンテーションの手法を用いる。この手法は特徴が一様である部分をピクセル単位で領域分割を行い、同時にカテゴリ認識も行う手法である。Badrinarayanan らの研究[2]は、畳み込みニューラルネットワーク (CNN) を応用して、セマンティックセグメンテーション用のネットワークを提案した。

竹木らの研究[3]は、セマンティックセグメンテーション手法[4]と鳥検出器の併用を行い、検出性能の向上を成功させた。提案手法で用いる手法ではないが、性能向上のアプローチとして共通すると考える。

3 ピクトグラム合成画像生成手法

提案手法の処理の流れを図1に示す。ピクトグラムの中でも人物が含まれるピクトグラムは万

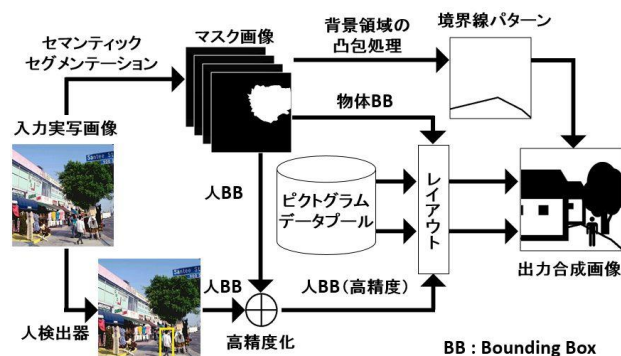


図1 提案手法の流れ

国共通で利用度が高いため、特に検出精度を上げる必要がある。そのため、提案手法では人検出器を追加する。

3.1 ピクトグラム画像

検出領域に合わせてピクトグラム画像を貼り付けるために、ピクトグラム画像を集めたデータベースを事前に準備する。ピクトグラム画像は透明度情報を持つPNG画像とする。

また、海や道など一部背景領域のピクトグラム画像は隣接領域との境界線に対応付ける。境界線パターンの生成には凸包を利用する。凸包は領域内を最小で囲む凸集合であり、各点に対して最短距離で描画されるため、ピクトグラムの単純化表現として適している。そのため、一部背景領域のピクトグラム画像はセマンティックセグメンテーションの後に、凸包を用いて作成する。

3.2 CNN

Badrinarayanan ら[2]を参考にセマンティックセグメンテーション用のCNNを構築する。提案手法で使うデータセットは、Tighe らの研究[4]で使用された SIFT-Flow データセットである。RGB画像と領域にラベリングされた画像の2488組の画像を使い、人や空など33クラスのセマンティックラベルに分類する。セグメンテーション結果で得たマスク画像はモルフォロジー処理で正規化する。

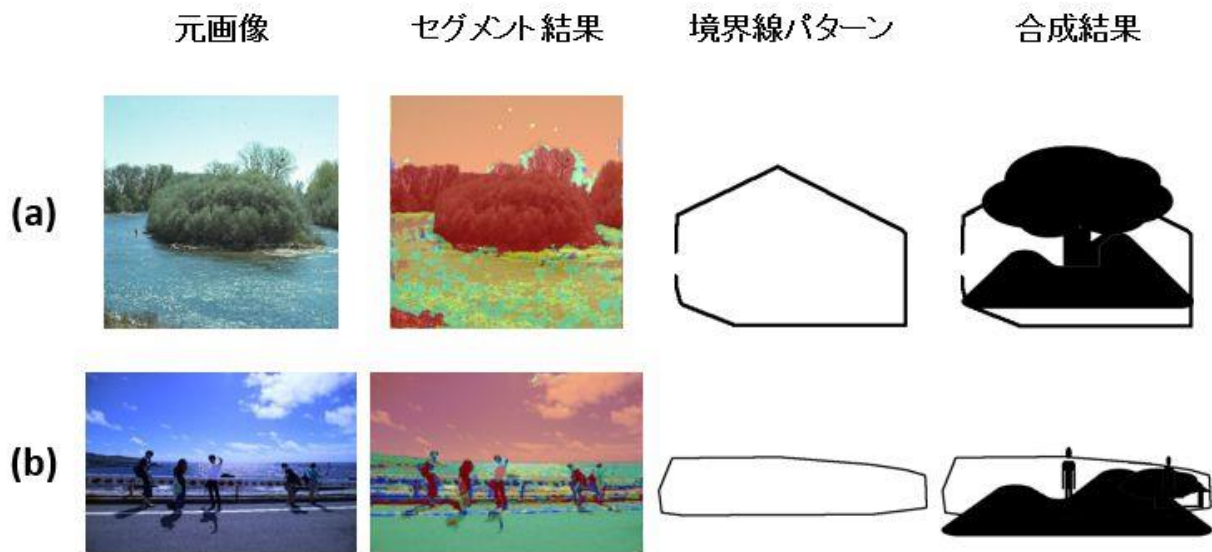


図 2 提案手法実装結果

3.3 人検出

提案手法では、MATLAB に搭載されている学習済みのサポートベクターマシン(SVM)を用いる。このSVMで検出するのは、陰に隠れていない直立姿勢の人物である。CNNの結果で得た人の検出領域と重複した場合、人検出器で得た位置を参照する。

4 初期実装結果

実写画像(a)と(b)を入力し、それぞれの処理結果を図2に示す。初期実装では検出した各領域の中でも一番大きな集合の部分に、検出した物体に該当するピクトグラムをデータプールから選出し配置する。大きさについては検出された領域のバウンディングボックスに合わせ、拡大縮小を行う。

4.1 実写画像(a)

セマンティックセグメンテーションで得たマスク画像を正規化した結果、おおよそ木 33%、空 31%、海 15%、山 12%、川 6%、道 3%の割合で領域を検出した。山領域は誤検出であり、主に水の領域部分で検出された。境界線パターンは空の部分で誤検出された画素を含んだことにより、画像の上部に伸びた。合成結果は誤検出により配置された山以外は妥当な結果になった。

4.2 実写画像(b)

(b)では、おおよそ空 49%、山 32%、建物 5%、木 5%、海 4%、川 3%の領域を検出した。(a)と同様に山領域が多く誤検出した。人の検出については、ほとんどが木の領域として誤検出し、正規化処理で消滅した。人検出器も、勾配情報が低い3人

の検出に失敗した。境界線パターンは空との区別をつけ、良い結果になった。合成結果は、誤検出の多さから良い結果にはならなかった。

5 おわりに

本研究は、CNNを用いたセマンティックセグメンテーションと人検出器からピクトグラム合成画像を生成した。初期実装の状態だが、静止画像に対してピクトグラムのような表現をすることができた。今後の課題は、CNNと検出器の精度向上、貼り付けるピクトグラムの個数や大きさの調整、人の身体方向やポーズに合わせたピクトグラムの追加である。これらによりクオリティの向上が期待できる。

参考文献

- [1] vwv11 | 富嶽三十六景,
<http://www.vwv11.com/works/36views-of-mt-fuji>
- [2] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation", arxiv:1511.00561, 2015
- [3] 竹木章人, チントウアントウ, 吉橋亮太, 川上玲, 飯田誠, 苗村健, "生態調査に向けた領域分割と検出器の組み合わせによる鳥画像検出", 映像情報メディア学会 2015 年年次大会予稿集
- [4] J. Tighe, and S. Lazebnik, "SuperParsing: Scalable Nonparametric Image Parsing with Superpixels", International Journal of Computer Vision, 2010