

物体認識における背景の有無による学習効果に関する一考察

Recognizing Objects Without a Back Ground Image Using Stereo Cameras.

津田 諒太 , 植村 渉

Ryota Tsuda and Wataru UEMURA

龍谷大学

Ryukoku University

tsuda@vega.elec.ryukoku.ac.jp, wataru@rins.ryukoku.ac.jp

Abstract

ファクトリーオートメーションを課題とした RoboCup リーグとして Logistics League がある。このリーグでは、工場内の加工機械を模した MPS と呼ばれる装置を用い、指示に応じて素材を加工する必要がある。MPS は高さ約 1m の直方体であり、上面は 70cm × 35cm であり、機械の種類によって上に載っている装置が異なる。ここで、上に載っている装置の違いから、MPS の種類を見分ける必要がある。現在は、MPS の長辺側の側面に ALVAR の AR タグを取り付けて、識別できるようにしているが、テクニカルチャレンジとしてマーカ無しでの識別が求められている。

本研究では、ニューラルネットワークを用いた学習を用いることで物体識別を行う。このとき、ステレオカメラの視差による距離情報を用いて、MPS の上のみを切り出すことで、識別率に与える影響を調べる。

1 はじめに

近年、ファクトリーオートメーションが注目されている。研究課題の一つとして工場における生産ラインの変化への適応があげられる[1]。このような工場では加工機械の位置や種類が頻繁に変化する。機器の情報を人の手で設定する代わりに、ロボット自身がカメラなどの入力機器を用いて加工機械の形状及び種類を識別する方法が必要とされている。

このような画像による物体識別には、画素マッチングや画像学習が用いられる。画素マッチングは入力画像と比較用画像内の同一座標の画素値が近似するかどうかで物体識別を行う。機械の種類ごとに大量の画像を保持する

必要があり、また明暗の変化に弱い。画像学習は対象の画像を多数用意し、画像内に共通する要素を見つけ出す。学習後はそれら要素のみを利用するため画像を保持しなくて良い。しかし学習に必要な画像は数千枚の規模であり、数千枚の画像の撮影には膨大な時間がかかる。工場での撮影を行うことは現実的でない。本研究では少ない枚数しか撮影できない環境を想定し、学習が効果的に進むように画像情報から学習に必要な部分のみを抜き出して用いることを考える。ここではカメラを 2 つ用意し、ステレオマッチングにより距離を推定することで、背景画像を除いた画像を用意しその学習効果について評価を行う。

2 ニューラルネットワーク

ニューラルネットワークとは脳構造を模した学習方法の一つであり、入力層、中間層、出力層に分かれている。層のそれぞれは前の層の出力に重み付けした総和を取り、その値によって自分が出力を行うかを決定する。この際の重みを変化させることで学習を行う。

今回のニューラルネットワークは多層パーセプトロンを用い、中間層は 1 層、学習アルゴリズムは逐次誤差逆伝搬アルゴリズムを用いる。ニューラルネットワークには複数のパラメータがあり、パラメータによって学習の良し悪しが変化する。今回は表 1 のパラメータで総組み合わせ (32 パターン) の学習を行う。

中間層数	100, 1000
最大繰り返し数	100, 1000
学習打ち切り最小誤差	0.000001, 0.00001
重み変更度合い	0.05, 0.1
重み変更抑制	0.05, 0.1

Table 1: ニューラルネットワークの設定パラメータ

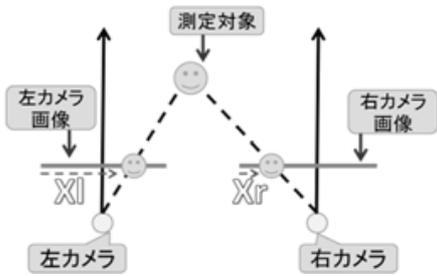


Figure 1: ステレオマッチングの原理

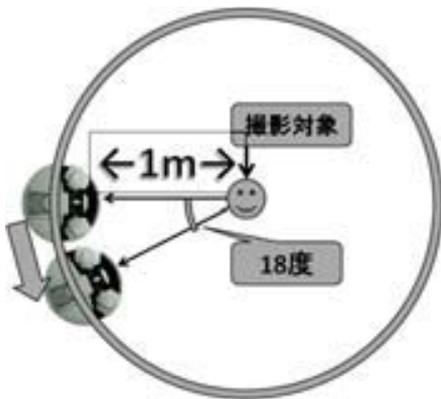


Figure 2: 撮影状況

3 ステレオマッチング

ステレオマッチングとは、2台のカメラを設置し、2枚の画像の視差を取ることで画像内の物体までの距離を推定する方法である。

基準となる左の画像内の要素に対応する要素を右の画像から探し、その視差 ($X_l - X_r$) から距離を算出する (図1)。

4 実験と結果

距離情報により背景を除いた場合の学習効果を確認するために、背景あり学習画像での学習と、背景なし学習画像での学習を行い、学習が収束した後に両者の誤認識率を比較する。

画像撮影は Festo 社から販売されている全方位移動可能ロボット Robotino を用いる (図2)。

Robotino は撮影対象から一定の距離を保ちながら対象の方向を向きながら移動し、撮影を行う。撮影は18度ごと、計20回撮影を行った。識別対象物として1) 持運びTV、2) スピーカ、3) コーヒーメーカーの3つを対象とした。背景あり画像は左カメラのものを使用した。背景なし画像は、左カメラ画像を基準に作成した距離画像の閾値以上の座標を元画像で黒く塗りつぶすことで背景を取り除いた。背景を除いた影響を測定するため、学習用画像と評価用画像の撮影は別の場所で行った。画像の明暗の影響

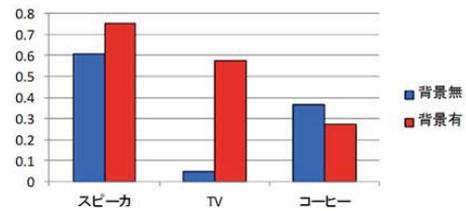


Figure 3: 背景の有無による誤認識率の違い

を排除するため、学習画像及び評価画像はエッジ抽出を行った。学習画像をニューラルネットワークに入力して、学習回数が最大繰り返しに達するか、学習結果の誤差が打ち切りの値以下になるまで学習を行った。完成した識別器に評価画像を入力し、識別結果を出力した。認識率と誤認識率の一部を図3に示す。

5 おわりに

大量の画像撮影ができない環境下において、少数の画像でも有効な学習結果を出力するために、ノイズとなりうる背景部分を除いての学習を提案した。背景有り画像での学習は、殆どの学習パラメータで有効な結果が出ず、最も優秀な結果だったものでも識別にはかなり不十分であった。一方、背景なし画像での学習は、背景なしと比較して優れた結果が出力された。複数枚の評価画像があれば十分識別可能な識別器が作成できた。今後の課題は、パラメータの設定に左右されない深層学習への適用が必要である。

謝辞

本研究は JSPS 科研費 15K16313 の助成を受けたものである。

参考文献

- [1] Frederik Zwillig, Tim Niemueller, Gerhard Lake-meyer, "Simulation for the RoboCup Logistics League with Real-World Environment Agency and Multi-level Abstraction", RoboCup 2014: Robot World Cup XVIII, pp 220 – 232, 2014.