

標的射撃競技者のマスク画像生成による特徴点選択手法の提案

	東海大学	*小柴 春樹	KOSHIBA Haruki
	東海大学	飯塚 泰樹	IIZUKA Yasuki
	東海大学	松本 哲志	MATSUMOTO Satoshi
01208380	防衛大学校	佐久間 大	SAKUMA Yutaka
02104890	東海大学	小林 正弘	KOBAYASHI Masahiro

1. はじめに

近年、多くのスポーツでコンピュータを利用した科学的トレーニングが注目されている。標的射撃もその1つである。標的射撃では、10m先の止まっている標的を狙う競技であり、トップレベルの選手は標的に半径5mm以下で集弾させており非常に精密な動作が要求される。そのため、フレーム間差分などによる従来手法の動き検知では、競技者の動きが小さすぎるために動きを検出することができない。したがって、競技者の動きを検出するためには競技者のわずかな動きを検出できるシステムが必要である。既存研究では、モーションキャプチャを用いた研究[1]などがある。しかし、これらの研究で用いられている機材は導入コストが高く、設置場所も限られてしまう。

本研究では、スマートフォンカメラで撮影した標的射撃映像から、競技者の動きを正確に抽出することを目標としている。既存手法には、映像から特徴点抽出とその動き検出を行い、検出された各特徴点の動きからフーリエ変換で取得した周波数的特徴をクラスタリングすることでノイズを除去する手法[2]がある。この手法では、競技者の上半身とライフルの特徴点を選択することが可能であるが、背景に映るわずかに動く特徴点を除去できないことやほとんど動きがない下半身を検出できない問題がある。しかし、射撃時の姿勢を正確に抽出するためには、競技者の全身とライフルについている特徴点をすべて抽出する必要がある。

本稿では、標的射撃映像からライフルを構えた競技者のみのマスク画像を生成する画像領域分割法とマスク領域内の特徴点を選択することで、解析に必要な特徴点の多くを失わずに高い精度でノイズを除去する特徴点選択手法を提案する。

2. マスク画像生成

画像領域分割法には、インスタンスセグメンテーションやセマンティックセグメンテーションのよう

に画像内の物体を認識し領域を分割する手法がある。しかし、画像に映っている物体すべてを認識してしまうため、対象の競技者の後ろに映っているほかの競技者の領域も抽出してしまう。

本手法では、pix2pixを用いて射撃映像から対象のライフルを構えている競技者のマスク画像を生成する手法を提案する。pix2pixとは、U-netを用いた敵対的生成ネットワーク(GAN)であり、入力画像と出力画像のペアを学習データとして読み込み対応関係を学習することで、1枚の画像から学習した対応関係に基づく補間を行った画像を生成することができる[3]。そのため、学習データにライフルを構えている競技者の画像とそのマスク画像を用いることで対象とするライフルを構えている競技者のマスク画像を生成することができる。

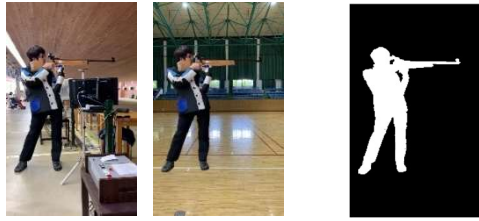
2.1 データセット

pix2pixでは、入力画像とその出力結果となる出力画像の画像データセットを用意する必要がある。そのため、競技者の画像とそのマスク画像を用意した。入力画像には射撃映像の最初のフレーム94枚と背景のみを変更した画像376枚を使用した(図1a)。このとき、背景を変更する画像には教室などの室内を中心にした背景用画像を18枚の中からランダムに4枚を用いて背景変更を行った。出力画像には、画像から手動で作成した競技者のマスク画像470枚を使用した(図1b)。また、入力画像と出力画像に対してデータ拡張を行いそれぞれ1880枚の画像を作成した。

2.2 モデルの学習

モデルの学習では、画像の前処理として画像のグレースケール化を行い、左右に余白を追加し画像を正方形にして256×256にリサイズした画像を使用して学習した。また、最適化アルゴリズムにAdamを使用しパラメータは $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ に設定した。

3. 学習したモデルを用いた特徴点選択



(a) 入力画像 (b) 出力画像

図 1 データセット画像

2 節で作成したデータセットを用いて学習した pix2pix の出力から以下の手順で特徴点選択を行う。

Step0 : 学習済みの pix2pix モデルに射撃動画の最初のフレームをグレースケール化し 256×256 にリサイズして入力する。

Step1 : 出力されたマスク画像にラベリング処理で画像の 2 値化を行い生成されたマスクの領域が最大の領域のみを取得することでノイズを除去し、競技者のマスク画像のみを取得する。

Step2 : 入力画像から特徴点抽出を行い、抽出された特徴点の座標のマスク画像の色を取得する。このとき、取得した色が白ならば特徴点として選択し、黒ならばノイズとなるので選択しない。

4. 比較実験

提案した手法を用いて特徴点選択の比較実験を行う。実験には、室内の射撃場にて人工照明下で標的射撃を行っている競技者 5 人の映像を使用した。また、評価計算のために抽出された特徴点を目視で選択し、正解データの作成を行った。

まず、マスク画像生成による特徴点選択手法について、epoch 数が 100, 500, 1000 で学習したモデルで実験を行い精度計算した結果を表 1 に示す。実験結果より epoch 数 100 が最も高い精度で特徴点を選択できていることが分かった。

表 1 epoch 数ごとの精度

epoch数	100	500	1000
正解率	96.4	96.3	95.7
適合率	89.7	89.3	88.8
再現率	98.6	98.9	97.2
特異度	95.5	95.2	95.1
f値	93.9	93.9	92.8

次に、既存手法のフーリエ変換による特徴点選択手法と提案したマスク画像生成による特徴点選択手法で精度の比較を行った結果を図 2 に示す。また、フーリエ変換による特徴点選択手法では、抽出した周波数的特徴に対して k-means で特徴点選択をした数値を用いた。

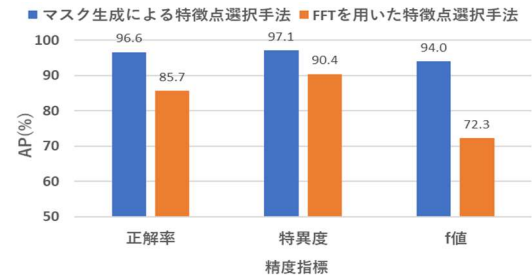


図 2 精度指標の比較

5. まとめ

実験結果より pix2pix を用いたマスク画像生成による特徴点選択手法はフーリエ変換を用いた特徴点選択手法と比較して正解率の値が平均 12%ほど高い精度で特徴点を選択できていることが分かった。また、f 値を見るとマスク画像を生成した場合ではフーリエ変換した場合より平均 20%程度高くなっている。これは、フーリエ変換した場合では下半身の動きがほとんどなかったために特徴点として選択できていなかったが、マスク画像を生成した手法では競技者が銃を構えているマスク画像を生成しているため下半身の特徴点をとらえることができたためだと考えられる。さらに、特異度の値は 97%程度となっており高い精度でノイズを除去できていると考えられる。

本稿では、スマートフォンカメラを用いて撮影した標的射撃映像から抽出された特徴点の選択手法について、マスク画像生成による特徴点選択手法を提案した。提案手法では、従来手法と比較して高い精度で特徴点を抽出できた。また、97%の精度でノイズを除去できていることが分かった。一方で、評価のための実験データ数が少ないため、今後はさらにデータ数を増やし比較実験を行う必要があると考えている。

参考文献

- [1] K. Mononen et al. "Relationships between postural balance, rifle stability and shooting accuracy among novice rifle shooters" Scandinavian Journal of Medicine & Science in Sports, vol.17(2), pp.180-185 (2006).
- [2] H. Koshiba et al. "A Video Analyzing Method for Competitive Rifle Shooting" Proc. 83th National Convention of IPSJ (2021).
- [3] P. Isola et al. "Image-to-Image Translation with Conditional Adversarial Networks" IEEE Conference on Computer Vision and Pattern Recognition (2017).