

複数視点からの観察による人姿勢認識のための視点計画

Viewpoint Planning for Human Pose Recognition from Multiple Viewpoint

○松田 耕太郎 (信州大学) Solvi Arnold (信州大学) 正 山崎 公俊 (信州大学)

Kotaro MATSUDA, Shinshu University 17w4059f@shinshu-u.ac.jp
Solvi ARNOLD, Shinshu University s_arnold@shinshu-u.ac.jp
Kimitoshi YAMAZAKI, Shinshu University kyamazaki@shinshu-u.ac.jp

In this paper, we proposed viewpoint planning method for human pose recognition who lying inside building. Taking multiple viewpoints is efficient to recognize complicated object like human poses. However, the minimum viewpoint sift is ideal. So, our method adopted already proposed framework witch choose efficient viewpoint based on probability calculation. The effectiveness of this framework has been demonstrated in several daily object. We made changes to this method by using new features extracted by convolutional neural network. This features obtain from middle layer the one before output layer of network have learned human posture. We evaluate this extracted features by compare with FPFH.

Key Words: Disaster Environment, Neural Network, Bayesian Inference, Viewpoint Feature

1. 緒言

近年, 災害対応の一策として, 災害環境に遠隔型のロボットを投入し救助活動に役立てようという動きがある. 人命救助において被災者を発見したとき, 対象の姿勢も同時にわかれば, 救助方法の決定に役立てることができる. しかし, 人が取りうる姿勢は多く, 単一の視点から見ただけではどのような姿勢かを確認することが難しい場合がある. そこで, 認識対象に関する情報を十分に得るために, 多視点からの観察を前提とし, 次に取るべき視点を決める視点計画が有用である.

視点計画については, 増沢ら[2]や松ヶ野ら[3]などいくつかの先行研究ある. 本稿では, 視点選択に確率の枠組みを取り入れた中里ら[1]の手法を導入し, 人の姿勢認識問題に対する有用性を調べる. 認識の信頼度を算出するための特徴量については, 3次元点群を入力とした特徴量である FPFH[4]を用いて, その有効性を調査する. また, 特徴抽出において, 認識汎用性の高いニューラルネットワークを用いて抽出する方法も試み, その有用性も調査する.

以降, 第2章では本研究で利用した視点計画の枠組みの概要を説明し, 第3章では今回導入したニューラルネットワークの特徴量と FPFH について述べる. さらに, 4章では検証に使用したデータと検証結果について述べる.

2. 視点計画の枠組み

2.1 視点計画の概要

視点計画には, 中里ら[1]の提案する多視点物体認識の枠組みを利用する. 視点計画と認識の流れを図1に沿って説明する. まず, 初期視点を取得後, 観測によって得られた点群データをもとにクラスごとの代表物体を選出する. 代表物体は事前に学習した物体データを探索し, 観測データに最も近いものが選出される. また学習データには視点情報が紐づけられており, 代表物体の選出によって推定初期視点も決まる. 次に分類するクラスの尤度比率を計算する. それぞれの代表物体は, 得られる特徴量の確率分布を持っている. それによって観測している対象が各クラスの代表物体に一致する尤度を計算し, 最終的にクラスの尤度比率とする. あるクラスの尤度比率が高いとき, 観測対象をそのクラスに属するものだと特定できたこととなり認識を終了とする. 一方で, クラスの尤度比率に差が見られない場合には, 観測対象を特定できるまでの特徴量が得られていない事を意味し, 追加視点の探索へと進む.

追加される視点は代表物体を利用して探索される. エントロピーと仮想観測データを用いて, 対象の認識に有効な視点を選択される. これを尤度比率が閾値を超えるまで繰り返す.

2.2 視点計画の数学的仕組み

エントロピーを用いた視点の選定方法と視点追加後の尤度比率の更新について説明する. エントロピーは観測物体の認識の不確かさを表し, 観測された情報量の程度を示す. エントロピー(平均情報量)は次式(1)によって定義される

$$H(P(X)) = -\sum_{N_x} P(X) \log P(X) \quad (1)$$

ここで, $P(x)$ は, 観測対象が x である確率であり, N_x は分類するクラスの数を表す.

視点の選定には, まず観測物体をあるクラスの代表物体だと仮定して移動先視点での特徴量を学習データから仮想的に取得する. この時, ある視点において得られる特徴量が仮定した物体に固有のものであればエントロピーは小さくなる. しかし, 各クラスの物体に一般にみられる特徴量は認識において情報量が十分でなく, エントロピーは大きくなる. このようにして学習データにある全視点を探索し, エントロピーが最

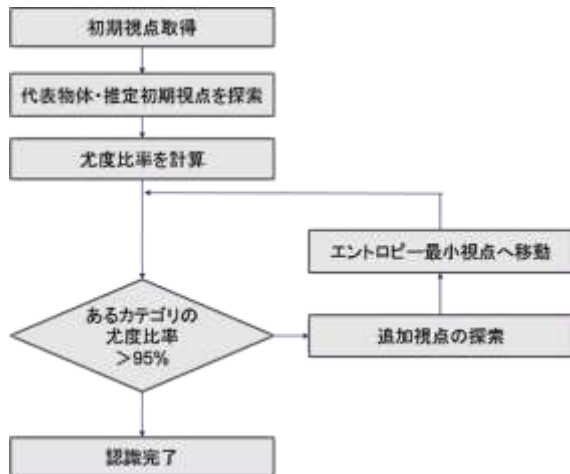


Fig. 1 Flow of recognition and viewpoint planning

小となる視点を選ぶことで効率的な視点の追加を実現している。

追加された視点では、再度尤度計算を行い各クラスに属する確率が計算される。ここで、前視点の確率を考慮した尤度比率とするため、ベイズの定理に基づいた事後確率を使用する。

$$p_{\text{post}}(c^{(n)}) = \frac{p(\mathbf{g}^{(n)} | c^{(n)}, \mathbf{v}^{(n)}) p_{\text{pri}}(c^{(n)})}{\sum_i p(\mathbf{g}^{(n)} | c_i^{(n)}, \mathbf{v}^{(n)}) p_{\text{pri}}(c_i^{(n)})} \quad (2)$$

ここで、 p_{post} ：視点取得後の事後確率、 p_{pri} ：視点追加前の事前確率、 c ：物体のクラス、 \mathbf{g} ：特徴量ベクトル、 \mathbf{v} ：視点ベクトル、添え字 n, i は n 視点目、クラス i を意味する。事前確率 p_{pri} は次の関係を満たす。

$$p_{\text{pri}}(c^n) = p_{\text{post}}(c^{n-1}) \quad (3)$$

3. 認識に用いる特徴量

2章で述べた視点計画では、物体の認識を左右する特徴量が重要な役割を果たしている。エントロピーを用いて効率的な追加視点を選定する際にも、特徴量がどれだけクラスごとに固有であるかが視点数削減のカギとなる。そこで本研究では、学習データによって柔軟に特徴量を抽出できるニューラルネットワークに注目し、新しい特徴量を使った視点計画を行った。

3.1 Fast Point Feature Histogram (FPFH)

はじめに中里らによって提案された手法[1]について実装を行い、人の姿勢認識問題に対する有用性を調べた。手法[1]では、視点計画に使用する特徴量として視点軸の回転に対して不変量である CVFH を用いていた。しかし、今回の問題設定では倒れている人の姿勢を認識対象とするため特徴量の回転に対する変化は重要である。そこで回転に対して変化し、且つ点群の基本的特徴量である FPFH を使用し有用性の検証を行った。さらに、FPFH を使った認識結果を基準として、本研究で提案するニューラルネットワークから得られた特徴量の有用性を検証した。

3.2 ニューラルネットワークから得られる特徴量

クラス分類を学習済みのニューラルネットワークにデータを入力すると、データの特徴付ける要素を抽出することが出来る。データに分類するクラスと同じ特徴がみられた場合にはネットワークの対応する部分が活性化し出力が変化する。本研究では、この出力の値を適切に取り出すことで特徴量として利用した。

まず、姿勢（人の姿勢）の深度画像を入力、姿勢の種類と視点を出力にとり、畳み込みニューラルネットワーク（CNN）を学習させる。視点は認識対象に対して周囲 360° を 5° 刻みで 72 点とした。三次元距離画像センサによる観測方法は図 2 に示すような条件で行い、三次元距離画像センサと中心点までの距離は 1m とした。図 2 中の θ は基準視点（頭側から見た視点）からの角度で現在の視点位置を表す。

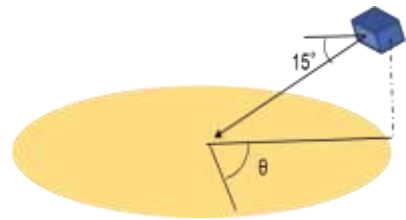


Fig.2 Viewpoint condition

畳み込みニューラルネットワークは、入力側から順に、畳み込み 3 層、全結合 1 層、出力層とした。出力層は並列に 2 つあり、姿勢の種類と視点位置をそれぞれの出力とした。姿勢の種類は、softmax 関数を使用した 3 姿勢の分類器となっている。視点位置の出力層は 2 つのノードから成り、座標点 $a(x, y)$ の値を出力する。原点から座標点 a までのベクトルと x 軸とのなす角により視点の位置を表現する。

このネットワークについて、出力層から一つ前層のノードの出力を特徴量として取り出す。次章の実験では、この層のノード数は経験的に 100 個とした。

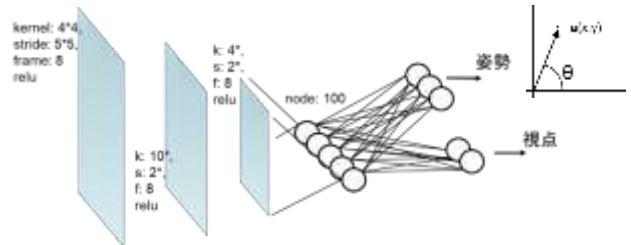


Fig.3 Structure of Convolutional Neural Network

4. 検証と考察

4.1 データの作成方法

姿勢認識の対象として 3 種類 {うつ伏せ・横向き寝・仰向け} を設定し、仮想環境でのデータ取得をおこなった。具体的には、ロボットシミュレーションによく用いられる Gazebo[5] の中で、SoftBank Robotics が開発した自立型ヒューマノイドロボット NAO[6] のモデルを上記 3 種類の姿勢で置き、Microsoft 社製の三次元距離画像センサ Kinect のモデルによって NAO モデルを仮想的に計測した。その出力から深度画像および 3 次元点群を得て、それぞれ FPFH の計算とニューラルネットワークの入力として使用した。3 種類の姿勢については、いずれも倒れている状態の頭部側の視点を角度 0° とした。

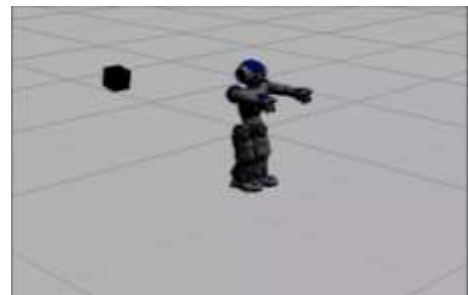


Fig.4 Gazebo simulator

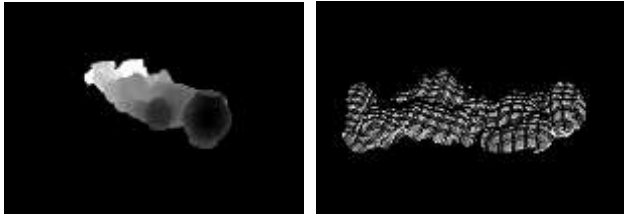


Fig.5 Depth image (left), Point cloud image (right)

4.2 検証結果

以下にそれぞれの方法で行った認識結果を示す。

Table 1 FPFH with random viewpoint

初期視点 \ 対象	うつ伏せ	横向き寝	仰向け
0°	0.983, 3	0.985, 3	0.008, 10 (うつ伏せ: 0.99)
45°	0.997, 6	0.981, 3	0.001, 20 (うつ伏せ: 0.98)
90°	0.987, 8	0.993, 5	0.0002, 21 (うつ伏せ: 0.99)
135	0.984, 5	0.986, 5	0.003, 16 (うつ伏せ: 0.99)
180	0.988, 5	0.984, 3	0.010, 6 (うつ伏せ: 0.98)

Table2 FPFH with viewpoint planning

初期視点 \ 対象	うつ伏せ	横向き寝	仰向け
0°	0.994, 3	0.992, 2	0.990, 5
45°	0.984, 2	0.991, 2	0.008, 4 (うつ伏せ: 0.98)
90°	0.994, 3	0.985, 2	0.008, 4 (うつ伏せ: 0.98)
135	0.990, 2	0.987, 2	0.009, 3 (うつ伏せ: 0.98)
180	0.989, 2	0.994, 2	0.991, 3

Table 3 CNN feature with viewpoint planning

初期視点 \ 対象	うつ伏せ	横向き寝	仰向け
0°	0.995, 3	0.989, 2	0.992, 3
45°	0.981, 2	0.990, 2	0.992, 3
90°	0.988, 2	0.991, 2	0.995, 3
135	0.985, 2	0.994, 2	0.995, 3
180	0.982, 2	0.992, 2	0.992, 3

表の列は、それぞれの姿勢を観測対象とした場合の認識結果(左)と取得した視点数(右)を載せている。表の行で、初期視点を 45° ごとに変えて検証を行った。結果が赤字で示されている部分は、認識に失敗しカッコ内の姿勢と誤認識していることを表す。

まず、手法[1]の有用性について考察する。特徴量に FPFH を用いてランダムに視点を取得した場合(表 1)と手法[1]により視点計画を行った場合(表 2)の認識結果を比較する。

仰向けにおいては、手法[1]を用いた視点計画の認識で視点数の大幅な減少がみられた。また、うつ伏せや横向き寝においても視点数が減少していることが確認された。これにより視点計画を用いる手法[1]による認識が人の姿勢認識においても有効であることが確認された。

次に、FPFH を用いた視点計画(表 2)と CNN から抽出された特徴量を用いた視点計画(表 3)の認識結果を比較する。FPFH では、仰向けの観察時に 45~135° から視点計画を開始した場合には、似た姿勢のうつ伏せと誤認識をしている。対して CNN から抽出された特徴量での認識結果は、どの初期視点においても観察対象を正しく認識している。また視点計画によって選択された視点の経路はそれぞれ特徴量ごとに異なり、CNN による学習で得られた特徴量が従来の FPFH とは違った要素を特徴量と見ていることも確認された。今後は CNN による特徴量が具体的にどのような部分の特徴量として学習しているかを明らかにしたい。

今回の検証においては、対象を 1m の距離から観察したため、3次元点群を使用した FPFH では点群の密度の影響があったことが考えられる。よって今後は、対象までの距離と認識性能を比較する必要があると考えられる。また、学習データが不十分であったり偏りがある場合などには、FPFH, CNN どちらの特徴量においても確率計算における分散が非常に小さくなり視点の追加が行われない場合が確認された。

5. 結言

本稿では、複数視点からの観察による人の姿勢認識のための視点計画を行った。従来の確率的な視点計画の枠組みに対して、物体を認識する新しい特徴量を用いてその性能を評価した。また検証の過程で学習データの偏りに対する視点計画の問題点も明らかになった。今後の課題としては、偏りの少ないデータの学習方法や分散に対してロバストな確率的手法の考案が挙げられる。

謝辞

本研究の一部は内閣府 ImPACT TRC の補助を受けた。

参考文献

- [1] 中里一幾, 諸岡健一, 辻徳生, 倉爪亮: “幾何的特徴量を用いた視点計画に基づく物体認識”, 第 32 回日本ロボット学会学術講演会予稿集, 1J1-5, 2014.
- [2] 増沢広朗, 三浦純, “移動ロボットによる時間制約を考慮した環境情報要約のための視点計画”, 日本ロボット学会誌, Vol.28, No.8, pp.938~945, 2010.
- [3] 松ヶ野修功, 高橋友和, 井出一郎, 日加田慶人, 村瀬洋: “複数視点物体認識における視点配置の決定法” 電子情報通信学会, pp.19-pp.24, 2004.
- [4] R. B. Rusu, N. Blodow and M. Beetz, "Fast Point Feature Histograms (FPFH) for 3D registration," 2009 IEEE International Conference on Robotics and Automation, Kobe, 2009, pp. 3212-3217.
- [5] <http://gazebosim.org/>
- [6] <https://www.ald.softbankrobotics.com/en/robots>