

Generative Adversarial Networks を用いた からあげピッキングにおける前処理としての画像変換

○今枝 航[†], 真野 航輔[†], 大島 彩佳里[‡], 平川 翼[†], 山下 隆義[†], 藤吉 弘亘[†]

[†]: 中部大学, [‡]: 三菱電機

wata@mprg.cs.chubu.ac.jp

概要：把持位置推定手法である Fast Graspability Evaluation は、デプス画像に対してハンドテンプレートを畳み込み、把持位置を推定することができる。しかし、実測データは抜けや値飛びといったノイズが多い。そのため、各ワークに把持位置推定を行うために、インスタンスセグメンテーション等の処理を行えない場合には把持位置の推定精度が低下すると考えられる。そこで、Generative Adversarial Networks の応用手法である SimGAN をベースとして、デプス画像を CG 画像に変換するノイズ除去を提案する。提案手法を用いることで、平滑化やバイラテラルフィルタなどのノイズ除去と比較して、より高精度な把持を確認した。

Generative Adversarial Networks, 画像変換, 食品ピッキング

1. はじめに

工場などに導入されているマニピュレーションロボットにおいて、物体の把持位置推定は重要なタスクである。把持位置推定の手法として Fast Graspability Evaluation(FGE)[1]が提案されている。FGE はデプス画像に対してハンドテンプレートなどを畳み込むことで、把持位置を推定することができる。しかし、実測データには抜けや値飛びなどのノイズが多い。そのため、各ワークに把持位置推定を行うために、インスタンスセグメンテーション等の処理を行えない場合には把持位置の推定精度が低下すると考えられる。そこで本研究では、深層学習における生成モデルの一つである Generative Adversarial Networks(GAN)[2]を用いて、ノイズを含まず処理が行いやすい CG 画像へ変換することにより、把持位置の推定精度の向上を図る。デプス画像の変換には、GAN の応用手法である SimGAN[3]をベースとした提案手法を用いて、ノイズを含まないCG画像へ変換する。これにより、デプス画像全体に対して FGE を用いた把持位置推定を行う際の推定精度の向上が期待される。

2. 関連研究

図 1 に示す FGE は、学習を必要とせずデプス画像より把持位置を推定する手法である。ロボットのハンドモデルより衝突領域と干渉領域のテンプレートを作成し、デプス画像に畳み込む。その出力に対してガウシアンフィルタを畳み込むことにより Graspability マップを作成する。この Graspability マップにおける最大値の位置を把持位置として検出する。また、ハンドテンプレートを作成する際のハンド開き幅や回転刻み角を

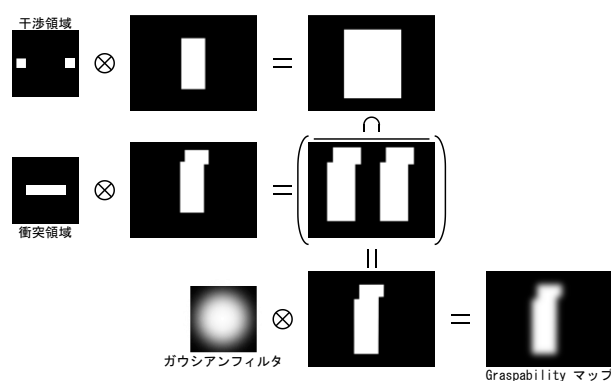


図 1 FGE における把持位置推定の流れ

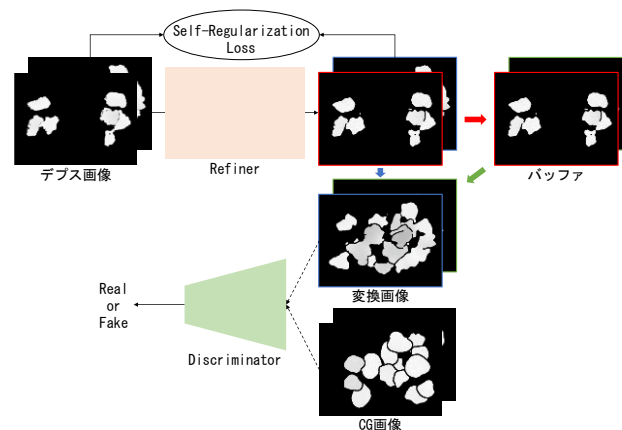


図 2 SimGAN の構造

調整することで、任意の分解能を得ることが出来る。

図 2 に示す SimGAN は、対応関係を持つペアの画像を変換する手法[4][5]とは異なり、ペアでない画像間における変換を行う手法である。また、SimGAN の Refiner は GAN における Generator である。変換元の画像が持つ特徴の変化を抑制するため、Self-Regularization loss を導入している。この損失関数は、変換前後の画像を写像関数 $f(\cdot)$ へ入力し、その画

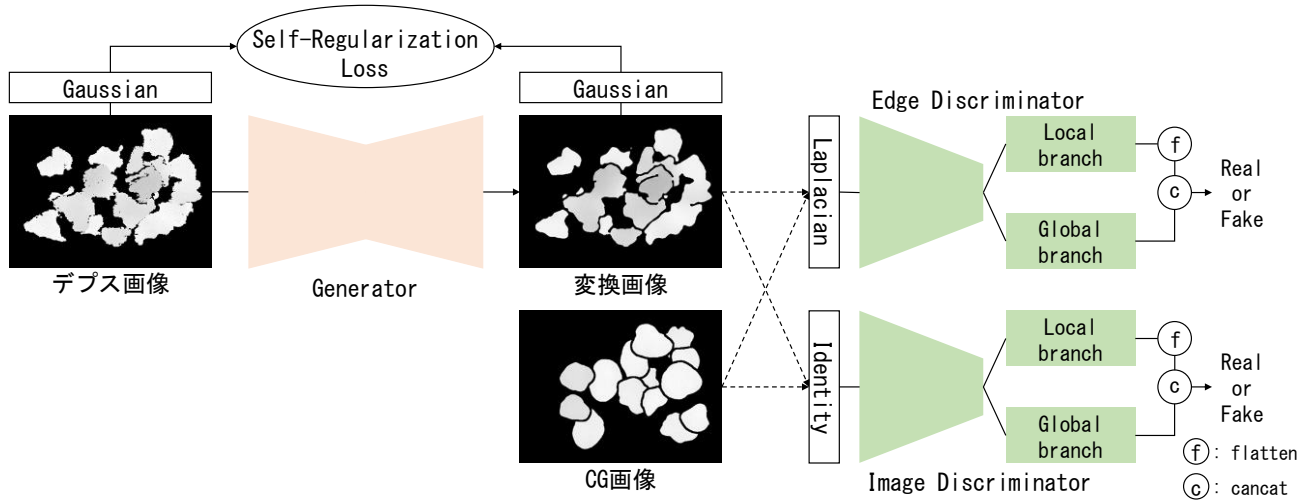


図 3 提案手法のネットワーク構造

像間の L1 ノルムより誤差を算出する. Discriminator では, 入力画像が実画像か生成画像かの識別をパッチごとで行う PatchGAN[4]を用いる. これにより, 細部まで高品質な変換ができる. また, 生成した画像の半分をバッファに保存し, 過去に生成した画像の一部を用いて Discriminator を学習することにより, 過去に生成したアーチファクトの再生成を抑制している.

3. 提案手法

本研究では, FGE に適した画像へ変換するために, SimGAN をベースとしたデプス画像の変換手法を提案する.

3.1. ネットワーク概要

図 3 にネットワークの概要を示す. デプス画像と CG 画像に生じるエッジの滑らかさやノイズの違いに着目する. Generator の学習において, Self-Regularization loss を用いて変換前後の画像による L1 ノルムでは, デプス画像が持つエッジやノイズの変化も抑制される. そこで, Self-Regularization loss における写像関数 $f(\cdot)$ にガウシアンフィルタを用いることで, エッジやノイズ以外の画像が持つ距離情報の変化を抑制できる. 次に, 通常の画像の識別を行う Image Discriminator とは別に, 入力を微分画像とする Edge Discriminator を追加する. これにより, CG 画像が持つ滑らかなエッジに注目した学習を行う.

3.2. Discriminator

提案手法の Edge Discriminator において入力に微分画像を用いた場合, パッチ画像に含まれる情報量は特に少なくなる. また, Image Discriminator においても変換対象であるワークの個数が少なくなったシーンでも同様である. そこで, Discriminator を各パッチに対して識別を行う Local branch と, 画像全体で識別

を行う Global branch に分割する. これにより, 情報の欠損を防ぎつつ, 局所的かつ大域的な特徴を捉えながら学習できる. ここで, Local branch は Receptive field が 96×96 ピクセルの PatchGAN の構造となっている.

3.3. 誤差関数設計

本手法で使用する敵対的誤差関数には, クロスエントロピー誤差関数を使用せずに LSGAN[6]の二乗誤差関数を用いる. そのため, Generator は式(1), Discriminator は式(2)を最小化することで各ネットワークの学習を行う.

$$L_g = \lambda_{edge} (D_e(G(x)))^2 + (1 - \lambda_{edge}) (D_i(G(x)))^2 + \lambda_{reg} \|f(G(x)) - f(x)\|_1 \quad (1)$$

$$L_d = \lambda_{edge} (D_e(y))^2 + (1 - \lambda_{edge}) (D_i(y))^2 + \lambda_{edge} (D_e(G(x)) - 1)^2 + (1 - \lambda_{edge}) (D_i(G(x)) - 1)^2 \quad (2)$$

ここで, x がデプス画像, y が CG 画像, G が Generator, D_i が Image Discriminator, D_e が Edge Discriminator である. また, 学習時のハイパーパラメータとして自己正則化強度の調整には λ_{reg} , 各 Discriminator における誤差のバランスを決定には λ_{edge} を用いて各誤差の調整を行う.

表 2 実機による把持の精度[%]

処理手法	把持成功率	物体落下率	複数把持率	把持失敗率
処理無し	88.0	6.0	0.0	6.0
ガウシアンフィルタ	91.0	6.0	1.0	2.0
バイラテラルフィルタ	87.0	4.0	0.0	9.0
提案手法	93.0	5.0	0.0	2.0

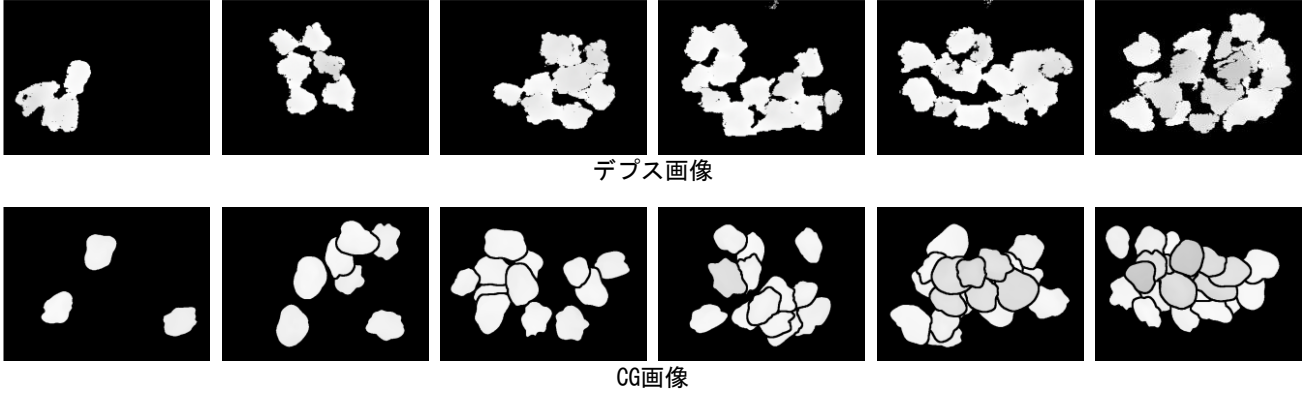


図 4 使用するデータセット

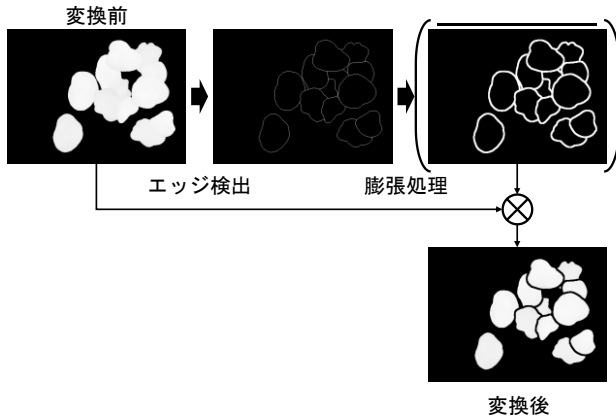


図 5 CG 画像の変換手順

4. 評価実験

提案手法の有効性を確認するために、変換前後の画像および、検出された把持位置の比較を行う。さらに、6 軸ロボットを用いた実機による把持実験を行い、変換画像を用いた際の把持成功率や物体落下率を比較し、その有効性を調査する

4.1. データセット

本実験で使用するデータセットには、把持実験で用いる実験環境にて撮影されたデプス画像と、からあげの 3D モデルを用いて作成した CG 画像を各 1,000 枚使用する。データセットの一部を図 4 に示す。ここで、1 枚の中には 1~20 個のワークを配置して撮影を行い、デプス画像のからあげは食品サンプルを用いている。また、実験環境で使用するデプスカメラには Intel RealSense Depth Camera D435 を用いる。CG 画



図 6 把持実験環境

表 1 把持実験における FGE のパラメータ

ハンドタイプ	2 指ハンド
開き幅	50, 55, 60, 65, 70[mm]
回転刻み角	5[deg]
進入深さ	42[mm]

像については、図 5 に示す手順にてワークの境界を付与した画像を用いている。

4.2. 実験概要

本実験では、画像サイズを 640×480 ピクセル、バッチサイズを 16 に設定し、最適化手法に Adam[7]を用いて、200 エポック学習する。学習時のデータ拡張には、垂直反転と水平反転を使用して画像の増幅を行う。誤差のハイパーパラメータである、自己正則化強度 λ_{reg} は 1、各 Discriminator における誤差のバランスを決定する λ_{edge} は 0.75 とする。また、提案手法で使用するガウシアンフィルタは 3×3 、ラプラシアンフィルタは 5×5 のカーネルを用いたフィルタを用いる。実

機実験は図 6 に示す環境にて行う。

4.3. 画像変換における定性的評価

提案手法による画像の変換結果を、図 7 に示す。図 7 より、計測が困難で滑らかなワーク境界付近において、提案手法で変換することにより CG 画像のような滑らかな形状に変換できていることが確認できる。また、一部の交換結果において、CG 画像に付与したワーク境界が生成されていることも確認できる。

4.3.1. 推定された把持位置の比較

FGE による把持位置推定結果を図 8 に示す。この時、比較対象として、ガウシアンフィルタやバイラテラルフィルタによるフィルタ処理を行った画像と比較する。また、FGE のパラメータを表 1 に示す。図 8 において、赤線で示されている把持位置が最終的に選択された把持位置である。提案手法で変換した画像の把持位置については、アプローチするワークの変化が確認できる。また、よりワーク中心に把持位置が推定できていることも確認できる。

4.4. 実機による把持実験

把持実験では、ばら積みされた食品サンプルのからあげに対して各手法に対して 100 回行う。実機による把持実験の結果を表 2 に示す。表 2 より、処理を行わない場合と比較して、提案手法を用いて変換することにより、把持成功率は 5.0 ポイントの向上、把持失敗率は 4.0 ポイントの削減に成功した。これは、図 8 で確認されたアプローチするワークの変化や、よりワークの中心に把持位置が推定できたためだと考えられる。バイラテラルフィルタに関しては、ガウシアンフィルタと比較して劣った結果となった。この結果に関して、バイラテラルフィルタはエッジ情報を残したばかりであるため、ワーク境界付近のノイズに対して効果的には働かず、変換無し程度の精度になったと考えられる。また、すべての結果において、把持失敗の要因に各ワークに対して得られたデプスが一回り小さいものがあり、そのワークに対して把持アプローチを行ったために把持位置が浅く把持できない場合が確認された。これらより、提案手法を用いてデプス画像を変換することは物体把持の前処理として有効である。

5. おわりに

本研究では、FGE を用いた把持検出において、セグメンテーションなどを用いることが出来ない場合における把持位置推定精度の向上を目的とした画像の

変換手法を提案した。画像変換による定性的評価では、提案手法を用いて変換することにより、ワーク境界のノイズに対して CG 画像のような滑らかな形状に変化し、把持位置推定においては、アプローチするワークの変化が確認できる。また、よりワーク中心に把持位置が推定できていることも確認できた。また、実機実験において、把持成功率は 5.0 ポイントの向上、把持失敗率は 4.0 ポイントの削減し、その有効性を確認した。今後は、より正確な物体境界を生成およびワークの補間をするために、カラー画像の導入を検討し、ネットワークなどもより適した構造を模索することによりさらなる精度向上を目指す。

参考文献

- [1] Y. Domae, H. Okuda, Y. Taguchi, K. Sumi, and T. Hirai “Fast Graspability Evaluation on Single Depth Maps for Bin Picking with General Grippers”, International Conference on Robotics and Automation, 2014.
- [2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. “Generative adversarial nets.” Neural Information Processing Systems, 2014.
- [3] A. Shrivastava, *et al.*, “Learning from Simulated and Unsupervised Images through Adversarial Training”, Computer Vision and Pattern Recognition, 2017.
- [4] P. Isola, J. Zhu, T. Zhou, and A. Efros. “Image-to-image translation with conditional adversarial nets.” Computer Vision and Pattern Recognition, 2017.
- [5] T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro. “High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs.”, Computer Vision and Pattern Recognition, 2018.
- [6] X. Mao, Q. Li, H. Xie, R. Lau, Z. Wang, and S. Smolley, “Least Squares Generative Adversarial Networks”, International Conference on Computer Vision, 2017.
- [7] D. Kingma, J. Ba, “Adam : A method for stochastic optimization,” arXiv:1412:6980, 2014.

今枝航：2019 年中部大学工学部ロボット理工学科卒業，現在同大学大学院工学研究科修士課程在学中。GAN を用いた画像変換ベースのドメイン適応の研究に従事。

真野航輔：2020 年中部大学大学院博士修士課程修了，現在同大学大学院博士後期課程在学中。深層学習を用いない把持位置推定手法における計算コストの削減に関する研究に従事。

大島彩佳里：2018 年北海道大学大学院情報科学研究科修士課程修了，現在三菱電機株式会社先端技術総合研究所に勤務。ロボットビジョン，ロボットの知能化に関する研究に従事。

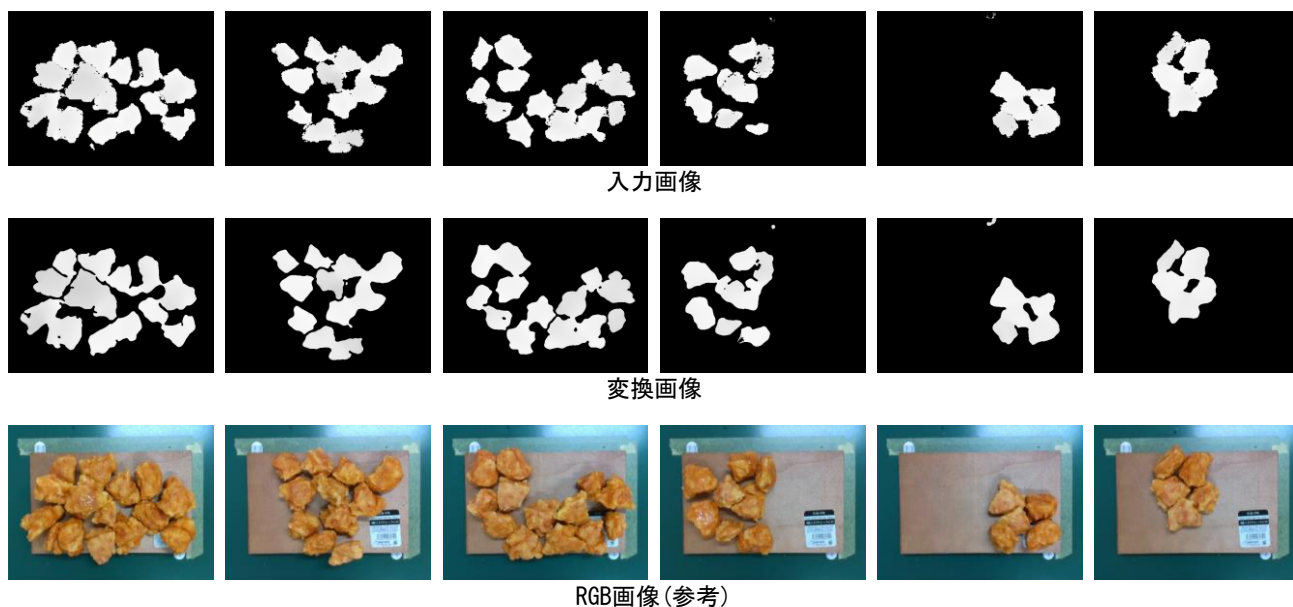


図 7 提案手法による変換結果

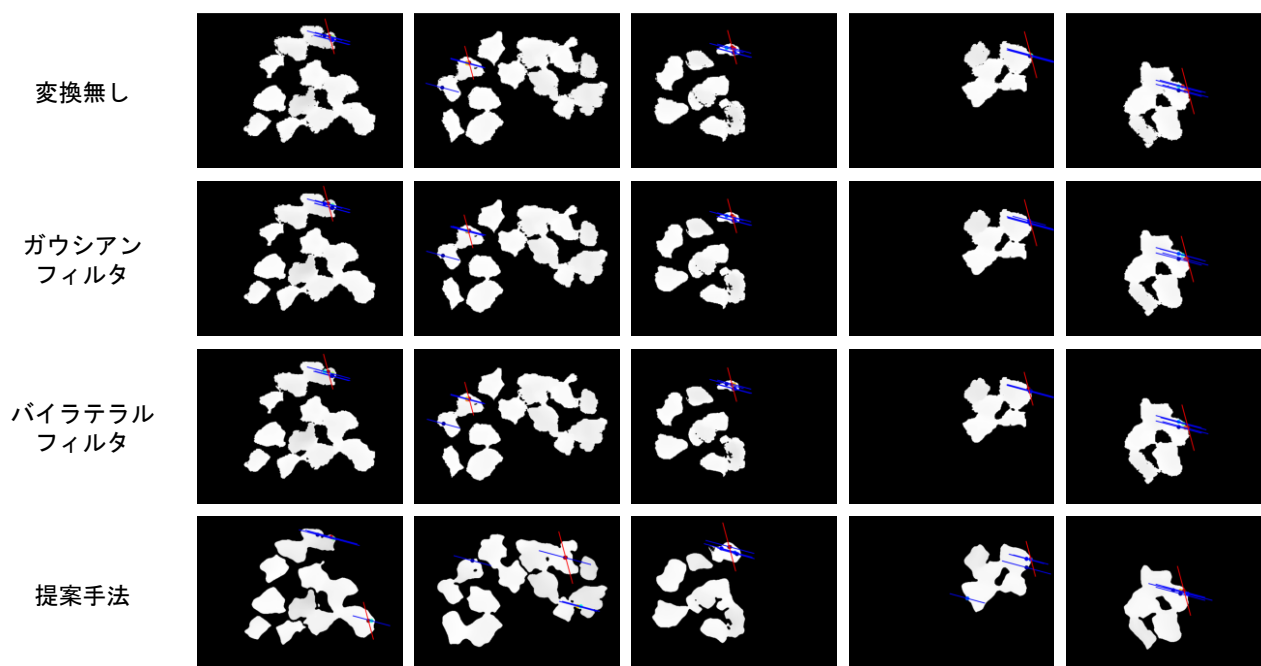


図 8 各変換手法による把持位置推定結果

平川翼：2013 年広島大学大学院博士前期課程修了，2014 年広島大学大学院博士課程後期入学，2017 年中部大学研究員（～2019 年），2017 年広島大学大学院博士後期課程修了．2019 年中部大学特任助教．2014 年独立行政法人日本学術振興会特別研究員 DC1．2014 年 ESIEE Paris 客員研究員（～2015 年）．コンピュータビジョン，パターン認識，医用画像処理の研究に従事．

山下隆義：2002 年奈良先端科学技術大学院大学大学院博士前期課程修了．2002 年オムロン株式会社入社，2009 年中部大学大学院博士後期課程修了（社会人ドクター），2014 年中部大学講師，2017 年より同大学准教授．人の理解に向けた動画処理，パターン認識・機械学習の研究に従事．

藤吉弘亘：1997 年中部大学大学院博士後期課程修了．1997～2000 年米カーネギーメロン大学ロボット工学研究所 Postdoctoral Fellow．2000 年中部大学講師．2004 年より同大学教授．2005～2006 年米カーネギーメロン大学ロボット工学研究所客員研究員，計算機視覚，動画処理，パターン認識・理解の研究に従事．