

Domain invariant な画像特徴による CG からの 能動ステレオシステムの学習

三嶋 道弘^{1,a)} 川崎 洋² 佐川 立昌³ 古川 亮¹

概要: 能動ステレオシステムを用いた三次元計測はその実用性により、様々な場面に用いられている。能動ステレオシステムは、プロジェクタとカメラから構成され、計測対象物体に特殊なパターンを投影しながら画像を取得する。そして取得した画像内の計測対象物体上に投影されたパターンの特徴点と、投影する元のパターンの特徴点との対応をとることで 3 次元点取得を行う。本論文では Domain invariant な画像特徴を用い、CG を使ったシミュレーション環境から得られた画像の特徴点と元パターンとの特徴点の対応関係を学習し、その学習結果を実環境から得られた画像に適用する。Domain invariant な特徴を用いることで、シミュレーション環境と実環境のドメインの違いを適合し、実画像でも精度よく対応点検出可能であることを示す。得られた対応点を用いて自動キャリブレーション、さらには、3 次元点取得可能であることを示す。

1. はじめに

本論文では、機械学習の際、学習データが含まれるドメインと実際に学習結果を適用する対象が含まれるドメインの違いから起こる学習効果の低下に対処するための手法を提案する。具体的には、能動ステレオシステムの対応点検出を目的とし、CG を使ったシミュレーションにより得られた画像上の特徴と元パターン上の特徴の対応関係を学習し、その学習結果を実環境から得られた画像に適用する。

能動ステレオシステムは三次元計測のために広く実用化されている。能動ステレオシステムでは、既知の情報を持つパターンを計測対象に投影する。一方で、パターンが投影された対象をカメラで撮影し、その画像を用いて復元を行う。その際、画像上の特徴点と投影パターン上の特徴点の対応関係を用いる。

能動ステレオシステムで使用される位置情報をエンコードするには、大きく分けて二つの方法がある。時間エンコードと空間エンコードである。前者は様々なパターン画像を変化させながら投影することで、一つのシーンを復元

する。後者は一種類のパターン画像を投影し、形状を復元する。そのため、人の動作の計測やロボットに搭載されるセンサなどに用いられることが多い。

実世界の三次元計測を行うには、医療現場や SLAM など、事前にシステムのキャリブレーションをできない状況が多い。そのような状況では、オートキャリブレーションが有効であるが、そのとき重要なのは対応点検出である。オートキャリブレーション可能な受動ステレオシステムでは、SfM といった手法が存在するが、それらは SIFT 特徴のような対応点のマッチングを利用している。

しかし、能動ステレオシステムでは、そのような対応点検出は難しい。これは、カメラ-プロジェクタシステムではエピポーラ拘束を仮定しているからである。また、特徴量を使った対応点検出は、投影する画像と投影された画像のドメインの違いにより、困難である。特に、回折光学素子を用いた DOE プロジェクタでは、パターンは固定されてしまうため、キャリブレーション用のパターン使用することができない。これらの理由によりキャリブレーションを自動で行うことは困難である。

格子パターンを用いた能動ステレオ手法は今までに多く提案されている [10], [13], [14]。そのためのパターンには光切断法に用いられ、さらに、格子点を用いた対応点推定に用いられる。近年、カメラとプロジェクタの位置を気にせず、手軽に三次元情報取得可能なシステムを構成するために、自動キャリブレーションの需要が高まっている。

本論文では、深層学習を用いて、カメラにより取得され

¹ 広島市立大学
Hiroshima City University, 3-4-1, Asaminami-ku, Oozukani-shi, Hiroshima-shi, Hiroshima 731-3194, Japan

² 九州大学
Kyushu University, 744, Motooka, Nishi-ku, Fukuoka-shi, Fukuoka 819-0395, Japan

³ 産業技術総合研究所
AIST, 1-1-1, Umezono, Tsukuba-shi, Ibaraki 305-8560, Japan

a) mikamo@hiroshima-cu.ac.jp

た計測対象上の投影パターンと、プロジェクタ投影する投影パターンの対応を効率的に検出する手法を提案する。さらに、学習によりえられた対応関係を実画像に適用することで、3次元計測可能なことを示す。具体的には、投影されたパターンと投影したパターンをグラフのマッチング問題として捉え、その対応関係を推定するために、Graph Convolutional Network (GCN) を用いる。提案手法では、さらに、システムのキャリブレーションから3次元計測までを単一のフレームワークで実現できる能動ステレオシステムを構築する。手法の主な貢献は以下の通りである。

- (1) GCN を用いたカメラ画像と投影パターンの対応検出と格子構造を持つ静的パターンを提案する。
- (2) 提案するシステムは GCN を用いることで、エピソード制約の有無に関わらず、プロジェクタとカメラの対応関係を推定することができる。その結果、既存のハードウェア構成を用いてシステムのキャリブレーションと3次元計測を実現することが可能である。

2. 関連研究

カメラ-プロジェクタシステムをキャリブレーションするために、様々な方法が提案されている [9], [15]。代表的な方法にはグレイコードのように時間的に符号化されたパターンを対象に投影する方法である。この方法では、複数のパターンを投影している間、投影対象は固定されている必要がある。また、プロジェクタとカメラのパラメータを推定するために、格子が印刷された平面にパターンを投影する方法もある [1], [12]。さらに、カメラで撮影される画像を合成することで、対応点を求める手法もある [15], [16]。これらは、キャリブレーションのための特別な物体や、複数の画像が必要である。

時間エンコードした構造化されたパターンを用いて、カメラ-プロジェクタ間の密な対応を得ることができる。それらの対応点を用いて、任意のシーンに対して自動キャリブレーションを適用することができる。それを利用して自己キャリブレーション可能な手法が提案されている [4], [8]。これらの方法では、対応点を検出しやすくするために、投影パターンをキャリブレーション用に特別に設計したり、もしくは、グレイコードのような時間的に符号化したパターンが利用される。しかし、DOE プロジェクタのような固定パターンしか投影できないプロジェクタでは利用できないという問題がある。

Furukawa らは内視鏡による体内形状計測のために、格子点の左右の高さに差を持つ静的パターン利用した自動キャリブレーション可能な手法を提案した。Furukawa らのパターンには9つのマーカーがあり、自動キャリブレーションのための情報に使用される。このパターンは 21×21 の格子点を持つ、密度の低い格子パターンを使用している。Nagamatu らは、格子状の構造を持ち、さらに、コードを

線と点の組み合わせで表現したパターンを提案した [10]。このパターンは 120×120 程度の格子点を持ち、高密度な測定が可能である。しかし、マーカーなどの情報がないため、エピソード制約がないと対応推定を予測することができないという問題がある。

カメラ-プロジェクタシステムを用いた能動ステレオシステムの場合、パターンの符号化方法は、時間的に符号化する方法と空間的に符号化する方法に分けられる。時間的に符号化する方法は複数の画像を必要とし、移動する物体の形状計測や DOE プロジェクタのような固定パターンプロジェクタを使用する場合には不向きである。また、空間的に符号化する手法の問題点には、キャプチャしたパターンが外乱によって再構成精度が大きく左右されることが挙げられる。最近ではこの問題に対処するために、取得した投影パターンを復号化するために、深層学習を用いることが提案されている [5], [6], [10]。

本論文では、コンピュータビジョン分野で広く利用されている Convolutional Neural Network (CNN) をベースとしたアルゴリズムと組み合わせることで、頑強に対応点検出可能な投影パターンを提案する。CNN には、ピクセル単位でラベル化された画像を生成するための FCNN アーキテクチャである U-Net [11] を [10] と同様に我々の目的に合うように修正した。Furukawa ら [7] は、内視鏡画像の撮影画像と投影パターンとの対応関係を推定するために GCN を用いた。この手法は、内視鏡画像の画像から検出されたパターンと投影パターンの対応関係を推定するために GCN を用いたネットワークが、直接対応点の番号を分類するクラス分けの問題を扱うのに対して、本研究では、画像から検出されたパターンと投影パターン間のノード間の類似度を出力する。このため、Furukawa らの手法よりも柔軟性が高い。

3. システムによる復元の流れ

提案手法では、能動ステレオシステムを用いた3次元復元に必要な対応点の検出をグラフのマッチング問題とみなし、これに Graph Convolutional Network (GCN) を用いる。図1にシステム全体の流れを示す。提案するシステム構成はキャリブレーションプロセスと3次元復元プロセスに分けることができる。まず、入力画像は、計測対象にパターンを投影したものをカメラで取得した画像である。この取得画像から、U-Net を利用して、線検出と各格子点の特徴の検出を行う。線検出により、画像中の格子点の位置情報を得ることができ、また、格子点の接続情報も得る。この格子点と特徴に対して GCN を利用したネットワークを適用し、対応点を検出する。その後、検出した対応点からキャリブレーションを行い、カメラ、プロジェクタの内部パラメータ、外部パラメータを算出する。算出したパラメータを用いて、再び得られた対応点を用いて3次元復元

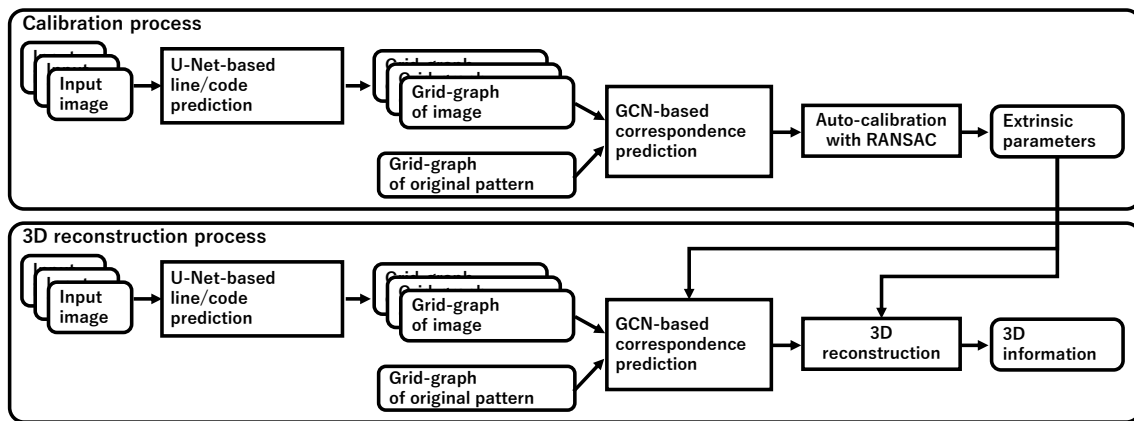


図 1 システムによる復元の流れ

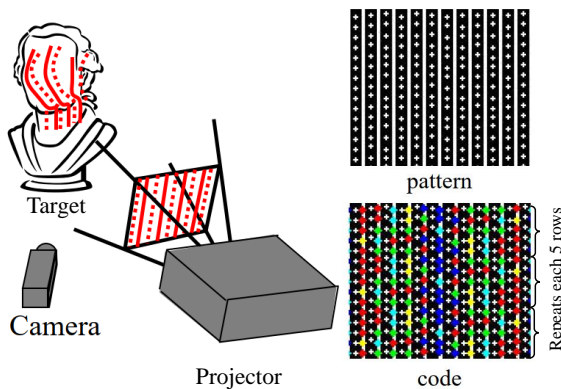


図 2 システムの構成と投影パターン.

を行う。

4. 実験環境

図 2 に提案するシステムの構成を示す。提案手法を適用するシステムはカメラとプロジェクタで構成された能動ステレオシステムである。特にプロジェクタは DOE プロジェクタであり、固定パターンを投影する。プロジェクタから固定パターンを投影し、その結果をカメラで撮影し、取得画像から形状復元を行う。固定パターンを使用するので、カメラとプロジェクタの同期は必要ない。DOE プロジェクタを用いるのは、想定する形状取得を行う環境に由来する。すなわち、システムを使用する環境を、散乱などが起こりやすい水中や、表面下散乱の起こりやすい体内など、一般的なプロジェクタでは投影パターンがはっきりと取得できない環境を想定している。そのような環境下ではレーザを用いることで、一般的なプロジェクタよりも散乱に対して頑強で、正確なパターンを取得できる。そのような理由から、提案手法では DOE プロジェクタを用いて多くの点が含まれたパターンを投影する。既存研究では Nagamatsu らは提案手法と同様な線と点を含むパターンを投影している [10]。手法では、検出の際、点をつなぎ合わせることで線とみなし、格子パターンを投影しているとみ

なし、処理を行っている。さらに、縦線は光切断法を用いた三次元形状計測に用いられる。提案手法でもこのフレームワークに従う。

図 2 の右側に、投影パターンと格子点の持つコードを色付けて示している。格子点は次のように決められる。まず、投影パターン中の十字の点は縦線をまたいで、つなぎ合わせることで線とみなす。そして、つなぎ合わせてできあがる線と、縦線との交点に格子点を設定する。

コードは格子点の左右の十字の点の位置関係、すなわち、格子点の左側の十字の点の位置に対して、格子点の右側のそれがどの位置にあるかにより、5 種類決まっている。この位置関係を示したのが図 3 である。図 3 中の黄色、赤、緑、青、水色はその 5 種類のコードを色で表している。すなわち、RR は右側が 2 段階高い位置にある、R は右側が 1 段階高い位置にある、S は左右同じ高さにある、L は右側が一段階低い位置にある、LL は右側が 2 段階低い位置にある、というように決めている。

また、図 2 中右下に示しているように、投影パターン全体に対して、コードは 5 行ごとに繰り返して配置してある。実際の投影パターンでは、縦方向に 120 個の格子点が配置されているため、 $120/5 = 24$ 回の繰り返しが存在する。復元の際は、この繰り返しに対して、エピソード拘束を使うことで一意に位置決定を行うことができる。しかし、キャリブレーションの際には、エピソード拘束は使えないため、この繰り返しを考慮したキャリブレーションが必要がある。各格子点の位置、番号、接続関係は図 4 に示す。

5. マルチタスク学習を用いた格子点とコード推定

提案手法では Nagamatsu らと同様にマルチタスク U-Net を用いて格子点の推定とその格子点の持つコードの推定を行う [10]。U-Net への入力は投影されたパターンを含む画像であり、U-Net からの出力は横線、縦線の検出結果、5 コードである (5)。横線、縦線の検出は格子点の位置情報を取得するために使用する。Nagamatsu らの U-Net の最

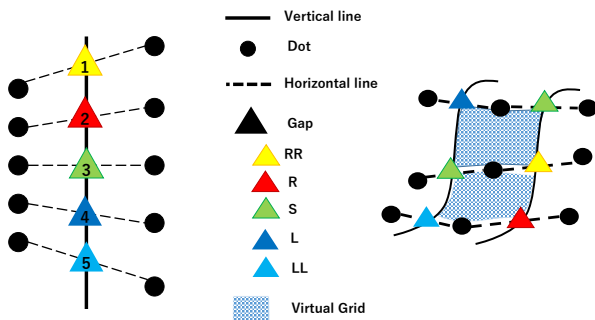


図 3 格子点を持つコードの種類

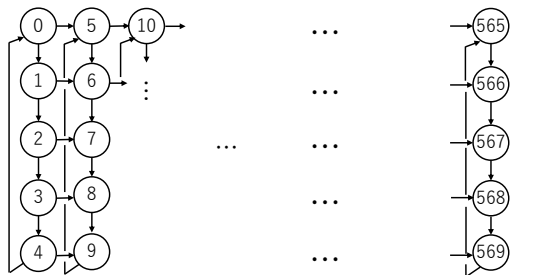


図 4 格子点の接続関係

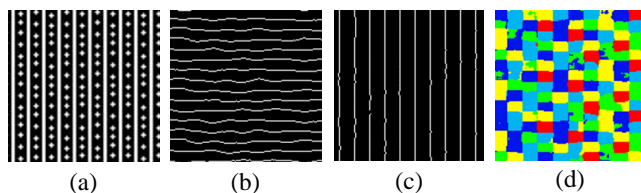


図 5 U-Net による推定結果

終層では、Softmax 関数を適用し、5 コードのうち最も大きい値を持つ特徴を、格子点を持つコードとして決定し、これを One-Hot ベクトルとして表した。例えば、格子点の右が一段階低いという情報を表す One-hot ベクトルは $[0\ 0\ 0\ 1\ 0]$ といった具合である。提案手法では、SoftMax 関数を適用する直前の値を正規化することで、5 コードのなかで最も大きい特徴だけを使用するのではなく、2 番目、3 番目といったすべてのコードについて情報を持つ特徴ベクトルを扱う。例えば、 $[0.00\ 0.00\ 0.05\ 0.52\ 0.43]$ のような特徴ベクトルである。

これにより、対応点を推定する際に最大確率を持つ格子点だけではなく、それに近い値を持つものも格子点の候補として推定することで、推定に柔軟性を持たせることを狙いとしている。

6. GCN を用いた対応点検出

6.1 GCN の利用

提案手法を用いた能動ステレオシステムでは、格子パターンを投影し、パターンが投影された対象を撮影した画像と投影したパターン画像から対応点を検出し、3 次元復元を行う。投影パターンの各格子点には隣り合う格子点が

存在する。この接続情報を利用することで、グラフのマッチング問題みなすことができる。そこで、提案手法では GCN を利用して、両者の対応関係を得る。

提案手法では取得画像から U-Net を用いて各格子点におけるコード情報と格子点の接続情報を取り出した後、投影パターン中の格子点を持つコード情報と格子点の接続情報とを比較することで、対応点を推定する。特に、提案手法では、GCN を利用したネットワークを特徴量を得るための変換器とみなす。すなわち、取得画像、投影パターンの画像それぞれから、ネットワークを通して特徴量を得た後、コサイン類似度を計算する。このコサイン類似度が大きくなるようにネットワークの学習を進める。このように学習することで、ネットワークは取得画像に対し、適切な特徴量を算出することができ、その変換を使って最も類似度の高い点を対応点として求める。

6.2 ネットワーク構造

提案するネットワーク構成を図 6 に示す。この図は取得画像に U-Net を適用し、得られた特徴ベクトルと接続情報を GCN を適用し、コサイン類似度を計算するための特徴量を算出するまでの流れを示している。これは次の手順で行う。提案手法では、格子の位置情報と格子のコード情報を格子の特徴として扱う。このため、格子点は 7 次元の特徴ベクトルを持つ。この 7 次元の特徴と接続情報を用いた対応点検出を実現するために、Graph Convolutional Network (GCN) [2] を用いる。GCN を適用することで、格子点のみの情報だけではなく、周囲の格子点を持つ情報も対応点を推定するために使用する。ネットワークの学習は、投影するパターンと投影されたパターンのコサイン類似度を最小にするように行う。このように学習し、U-Net から得られた特徴と接続情報を入力として、対応点の類似度を出力するネットワークを得る。

まず、U-Net により、縦線と横線、格子点の位置、各格子点の特徴ベクトルを得る。また、U-Net から得られる接続情報も用いて、これらの格子点のグラフ情報を提案する GCN を用いたネットワークへの入力とする。図 6 中の $H^{(0)}$ は U-Net により得られた特徴ベクトル、 $A_0 \sim A_3$ はそれぞれ各格子点の上下左右の接続情報を表す。提案するネットワークでは、これらの入力に対して、グラフコンボリューション、バッチノーマライゼーション、ReLU 層を 5 回繰り返した後、さらに Linear、ReLU 層を適用して、最終的な特徴を得る。

図 7 にコサイン類似度を計算する過程を示す。取得画像から U-Net を用いて得られるグラフを G_d 、また、投影パターンから得られるグラフを G_p とする。取得画像から得られたグラフと投影パターン中のグラフのそれぞれに対し、格子点の数が N_d と N_p であったとする。このとき、ネットワークへの入力は $N_d \times 7$ 、 $N_p \times 7$ の行列である。ネッ

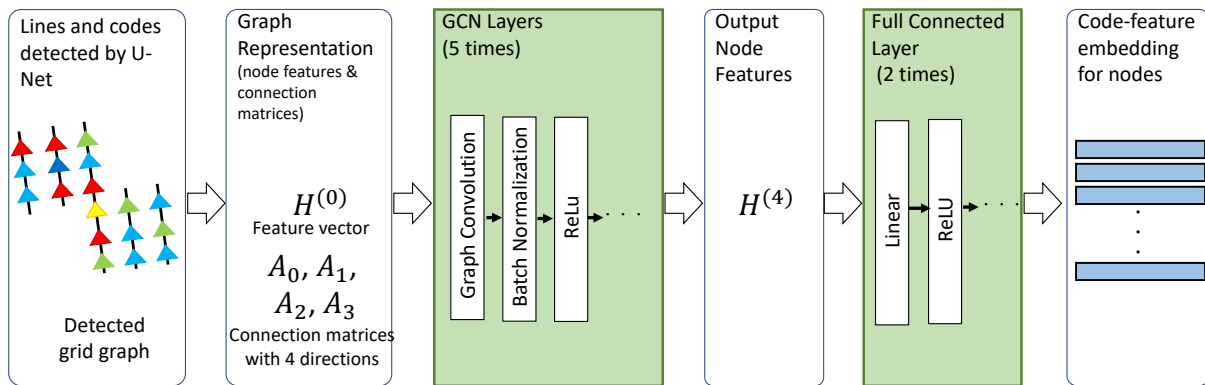


図 6 提案するネットワークの構造

トワークでは GCN を適用することにより、さらに特徴は増やされ、取得画像の特徴量を表す行列 \mathbf{A} とパターン画像の特徴量を表す行列 \mathbf{B} が得られる。このとき行列のサイズは $\mathbf{N}_d \times D_e$, $\mathbf{N}_p \times D_e$ となる。これらからコサイン類似度を計算し、類似度が大きくなるように学習を行う。これにより、対応点の類似度を推定する学習が可能になる。ここで類似度が高い点ほど、対応する点であることを意味している。

7. 繰り返しパターンを含む場合の対応点のキャリブレーション

前節で得られた G_d から G_p への対応は、元パターンの繰り返しグループへの対応であり、対応する格子点は一意に決めることができない。これを解決するために、隣接ノードどうしの対応の連続性を利用する。具体的には対応の繰り返しについて、位相アンラップし、最終的に一意に決める。パターンの繰り返しは最大で $120/5 = 24$ なので、それぞれでキャリブレーションを試み、RANSAC を用いて最も当てはまりの良い解を選択する。

8. エピポーラ拘束を用いた対応点検出

カメラとプロジェクタのキャリブレーションを行った後、提案手法では、エピポーラ拘束による重みづけと格子点の類似度を用いることで、用いない場合よりも、精度よく格子点探索する。計算された格子点の類似度とは、ある格子点に対し、全ての格子点が、それぞれの程度類似しているかを表す。また、能動ステレオシステムを利用し、エピポーラ拘束を用いることで、着目する格子点が投影パターン上のどの点に対応するのか、探索範囲を制限する。提案手法では、この特性を利用し、エピポーラ線に近いほど重みが大きくなるように設定する。この格子点の類似度とエピポーラ線による重みを重ね合わせて用いることで、一意に格子点に対応付ける。

この流れは次式を用いて表すことができる。格子点の類似度を表した行列を $\mathbf{M}_s \equiv \mathbf{AB}^T$ とする。提案するネットワークにより得られた類似度行列 \mathbf{M}_e は、投影パターン

全体を構成する繰り返し部分についてだけである。そのため、繰り返しの数だけコピーすることで投影パターン全体に対する類似度行列 \mathbf{M}_s^o を得る。今、着目する格子点をパターン上に投影したとき、その格子点 j とエピポーラ線とそのほかの格子点 i の距離を $Dist(i, j)$ とし、その重みを次式で定義する。

$$\mathbf{M}_e^o \equiv -\max(Dist(i, j) - T_e, 0)^2 \quad (1)$$

ここで T_e はマージンであり、 $Dist(i, j)$ が T_e よりも小さい場合、エピポーラ拘束による重みは 0 になる。最終的に対応点検出に使用する類似度は、この重みと GCN による類似度を足し合わせる。

$$\mathbf{M}_t^o \equiv \mathbf{M}_s^o + w_e \mathbf{M}_e^o \quad (2)$$

ここで w_e はエピポーラ拘束による重みづけを適用する場合の重みである。この重みを 0 にすることで、エピポーラ拘束による重みづけを行うか、行わないかを切り替えることができる。

9. 実験結果

9.1 CG を用いた学習データの作成

深層学習を適用するためには、十分な学習用のデータセットが必要である。提案手法では、線と点を含む特殊なパターンを投影するため、実際にシステムを構成して学習用のデータセットを大量に集めることは困難である。そのため、我々は CG を用いて学習用のデータセットを作成した。

任意の表面形状にパターンを投影した場合の学習データを大量に作成するために、プロジェクションマッピングをシミュレーション可能なアプリケーションを開発した。また、パターンだけではなく、格子点の番号を色に対応させた画像も投影することで、パターンが投影された画像と、それぞれの格子点の番号との対応関係を得た。図 8 にその一例を示す。左にアプリケーションを用いて投影した形状、中央には、ドットラインパターンを投影して得られる画像、右には格子点の番号を色付けして投影し、得られる

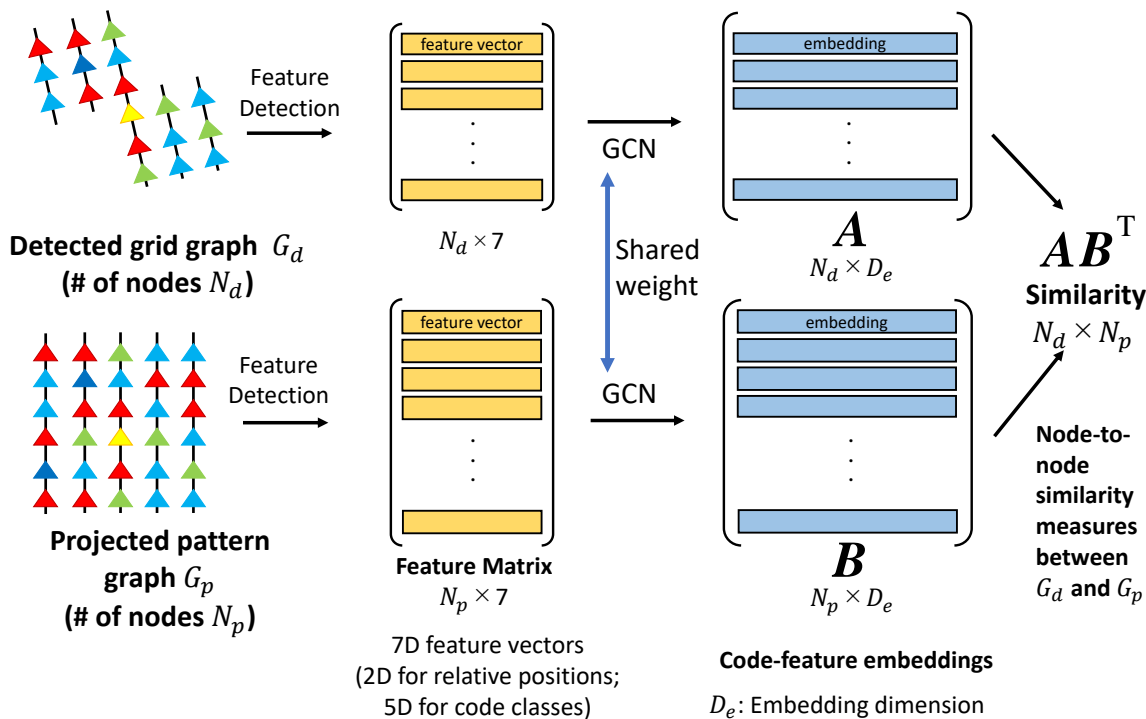


図7 類似度の計算の流れ

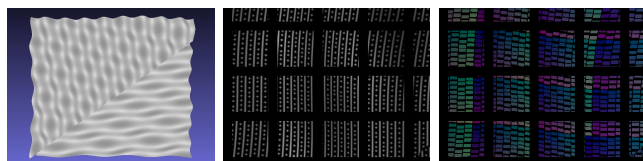


図8 学習画像取得のために作成した形状(左), CGを使ったパターン投影により得た画像(中央), 対応する格子点番号を色付けた画像(右)

画像を示した。教師データには中央の画像を U-Net に適用して得られる格子点の位置と特徴ベクトルと、その位置情報をもとにして右側の画像から得られる格子点の番号の組を用いた。

様々な表面形状の物体に対して学習を行うため、様々な波長をもつサイン波の形状を利用した。さらに、実際の形状にパターンを投影したときに、ドットラインパターンが不連続になる場合を想定し、二つの波面を組み合わせるように配置した。波面一つの方程式は次式で表される。

$$f(x, y) = A \sin(a(x \cos \theta_0 + y \sin \theta_0)) \quad (3)$$

このとき、振幅 A や波長を変えるために a を変化させて波面の形状を作成した。さらに、図8に示すように学習画像を格子状に分けることで、実際の形状にパターン投影した場合に不連続な領域が細切れになるような場合も想定している。

9.2 評価と復元結果

提案手法を評価するための実験を行った。まず、提案するネットワークによる対応点検出と、そのほかの対応点検

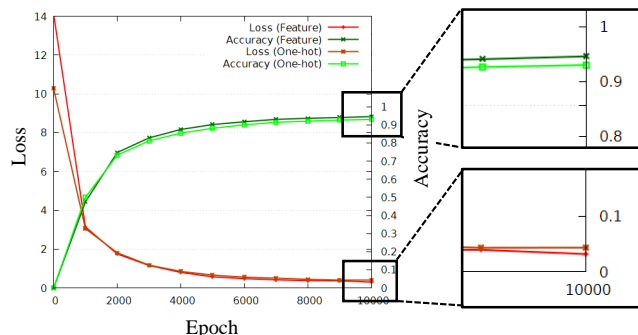


図9 One-hot ベクトルを用いた場合と特徴ベクトルを用いた場合の学習過程の違い

出手法として、Loopy Belief Propagation (LBP)[3]を用いた場合、精度を比較した。さらに、提案するネットワークの入力に One-hot ベクトルを持つ場合と特徴ベクトルを持つ場合を比較した。

図9にそれぞれ学習時における Loss と正解率の推移を示した。Loss(Feature), Accuracy(Feature) がそれぞれ、特徴ベクトルの Loss と正解率、また、Loss(One-hot), Accuracy(One-hot) が One-hot ベクトルにおける Loss と正解率を表す。正解率は、それぞれ、学習データを用いて、U-Net を適用して得られた特徴ベクトル、One-hot ベクトルと接続情報を適用した結果を GCN に適用した結果、格子点の番号が正しく推定できた割合を表している。図9からわかるように、学習が進むにつれ、特徴ベクトルが One-hot ベクトルよりも Loss が小さくまた、正解率も高くなるのがわかる。

次に、提案手法による対応点検出結果が他の手法よりも

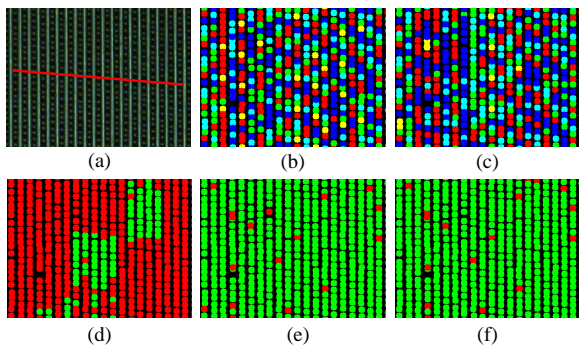


図 10 回転した画像に対する LBP と GCN の正解率の比較. (a) 評価に用いた画像. 赤線の傾きが回転量を表す. (b) 回転させる前の画像から得られる格子のコード情報. 5色で塗り分けられた各格子点の色はコードの種類を表す. (c) 回転後の格子点のコード情報. (d)LBP を用いて正解率を可視化した結果. 赤の格子点は誤って番号を推定した格子点, 緑の格子点は正しく番号を推定した格子点を示す. (e)GCN(One-hot ベクトル) を用いた場合の正解格子点を可視化した結果. (f)GCN(特徴ベクトル) を用いた場合の正解格子点を可視化した結果.

優れていることを示す. 図 10 に提案するネットワークを使った場合の対応点検出結果と, 比較対象として LBP を用いた対応点検出の結果を示す. この実験の目的は, 回転が含まれる画像に対して, 適切に番号を推定できるかどうかを調べることである. (a) に評価に使用した入力画像を示す. この画像は U-Net によるコードの判定に間違いが多く含まれるように, 我々が意図的に回転させたものである. 回転前の画像に U-Net を適用した場合に得られるコードを色付けて (b) に示す. 回転後と比較のため, 格子点が図中 (c) と同じ位置になるように配置している. 回転後の画像に U-Net を適用した場合に得られたコードを色付けて (c) に示す. (c) では, (b) と比べて, いくらの格子点の色付けが異なっている. これは, U-Net が (b) と (c) で異なるコードであると判断した結果である. すなわち, 回転により, 格子点の左右の高さが変化したためである. (d), (e), (f) は (c) に対して, それぞれ LBP, GCN(One-hot ベクトル), GCN(特徴ベクトル) を適用したときの, 正誤を可視化したものである. すなわち, 格子点が緑色の場合, 推定された番号と教師の番号が同じであったことを意味し, 赤の場合, それが間違っていたことを意味する. このとき, 正誤とは (c) のコードを入力として, (b) のコードを持つ格子点が推定されるかを示した結果である. (d), (e), (f) では, 提案するネットワークは LBP と比べて頑強であることを示しており, 正解率はそれぞれ 0.320, 0.925, 0.928 である.

図 11 にエピソード拘束を用いる前と用いた後の対応点推定とその復元結果の違いを示す. (a) はパターンを投影して取得した画像, (b) はエピソード拘束を用いずに対応点推定を行った結果, (c) はエピソード拘束を用いて対応

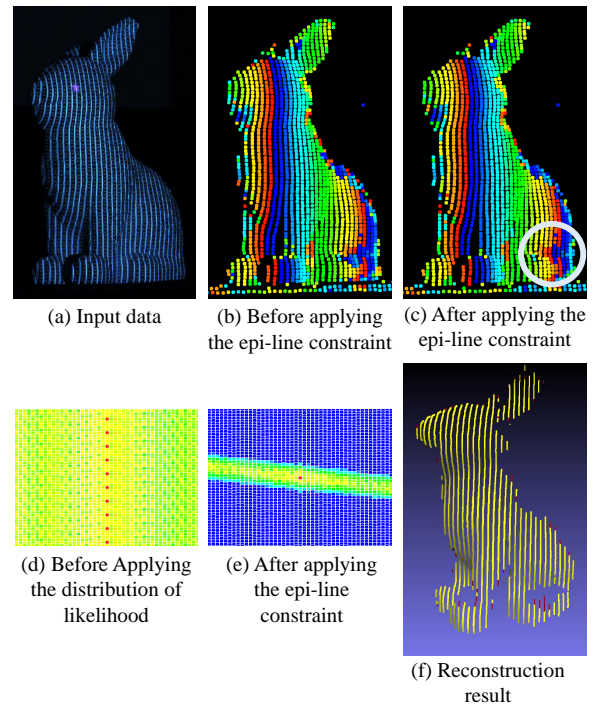


図 11 エピソード拘束を用いる前と用いた後の対応点推定と復元結果の違い. (f) は復元結果を表し, 黄色い線はエピソード拘束を用いていない場合の復元結果, それに加えて, エピソード拘束を用いることで, さらに赤線のように復元可能な領域が増えることを示している.

点推定を行った結果, (d) は格子点のうちの一つの点との類似度を表示した結果, (e) は着目する格子点に対して, エピソード拘束を用いて類似度を再計算した結果, (f) が復元結果である. 類似度は赤いほど高いことを表す. (b) と (c) 中の縦に同一な色は, 同じ列の格子点を表している. (b) から (c) にかけて, 白丸で囲んだところを見ると, 検出された対応点が増えていることがわかる. また, (d) では繰り返しが含まれている類似度から, (e) では, エピソード拘束を用いることで, その繰り返しがなくなり, 一意に決まっていることがわかる. さらに, (f) では, 実際に3次元復元を行った場合, 検出された対応点によって復元された領域が増えていることがわかる.

図 12 に提案手法を用いて復元した結果を示す. 提案手法では DOE プロジェクタを用いて計測対象にパターンを投影した. DOE プロジェクタは投影パターンを変えることができず, そのキャリブレーションは一般的な手法では難しい. 提案手法では自己キャリブレーションが可能であり, 図 12 に示すような凹凸を持つ三次元物体を復元することができる.

10. まとめと今後の課題

この論文では, Domain invariant な画像特徴を用いた能動ステレオシステムの学習方法を提案した. 提案手法では, 対応関係を得る問題をグラフのマッチング問題とみなすこ

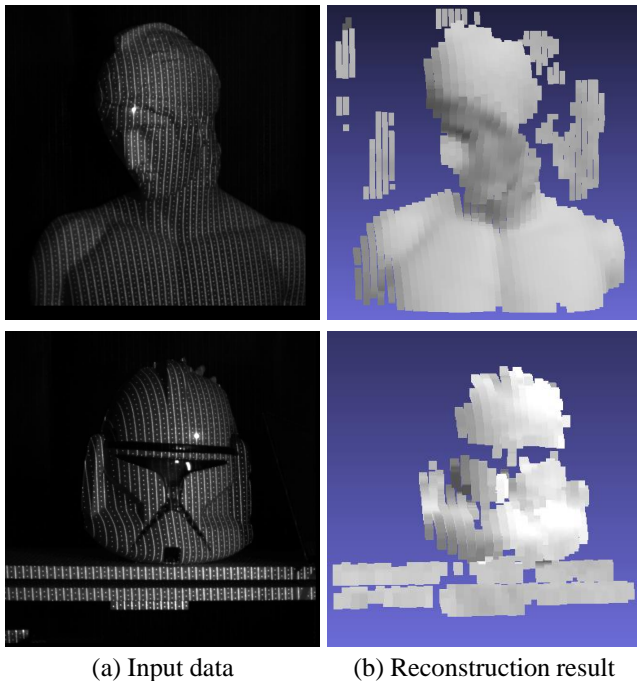


図 12 DOE プロジェクタを用いた復元結果

とで, GCN を利用する. 学習では CG により得られた画像を教師画像として用いて, 投影パターンとの対応関係を学習する. まず, CG により得られた画像に対して U-Net を使用して格子点と特徴ベクトル, 接続情報を取得する. それらを入力として, GCN を使用して対応点を検出する. 検出された対応点は, システムのキャリブレーションを行うために使用する. さらに, キャリブレーション後にはエビポーラ拘束を適用することで, さらに対応点検出の精度を上げる. 提案手法を構成するシステムは DOE プロジェクタのような固定パターンを投影するプロジェクタを仮定しており, 時間変化するパターンを投影できない状況でのキャリブレーション方法を提案した. 今後の課題としては, 投影パターンの工夫などにより更なる検出可能な対応点を増やし, より細かな形状復元を可能にすることなどが挙げられる.

参考文献

- [1] Audet, S. and Okutomi, M.: A user-friendly method to geometrically calibrate projector-camera systems, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, IEEE, pp. 47–54 (2009).
- [2] Defferrard, M., Bresson, X. and Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering, *Advances in neural information processing systems*, pp. 3844–3852 (2016).
- [3] Felzenszwalb, P. F. and Huttenlocher, D. P.: Efficient belief propagation for early vision, *International journal of computer vision*, Vol. 70, No. 1, pp. 41–54 (2006).
- [4] Furukawa, R. and Kawasaki, H.: Uncalibrated multiple image stereo system with arbitrarily movable camera and projector for wide range scanning, *Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM'05)*, IEEE, pp. 302–309 (2005).

- [5] Furukawa, R., Miyazaki, D., Baba, M., Hiura, S. and Kawasaki, H.: Robust structured light system against subsurface scattering effects achieved by CNN-based pattern detection and decoding algorithm, *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 0–0 (2018).
- [6] Furukawa, R., Morinaga, H., Sanomura, Y., Tanaka, S., Yoshida, S. and Kawasaki, H.: Shape acquisition and registration for 3D endoscope based on grid pattern projection, *European Conference on Computer Vision*, Springer, pp. 399–415 (2016).
- [7] Furukawa, R., Oka, S., Kotachi, T., Okamoto, Y., Tanaka, S., Sagawa, R. and Kawasaki, H.: Fully Auto-calibrated Active-stereo-based 3D Endoscopic System using Correspondence Estimation with Graph Convolutional Network, *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, pp. 4357–4360 (2020).
- [8] Li, C., Monno, Y., Hidaka, H. and Okutomi, M.: Pro-Cam SSfM: Projector-Camera System for Structure and Spectral Reflectance from Motion, *Proceedings of the International Conference on Computer Vision (ICCV)*, pp. 2414–2423 (2019).
- [9] Moreno, D. and Taubin, G.: Simple, accurate, and robust projector-camera calibration, *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, IEEE, pp. 464–471 (2012).
- [10] Nagamatsu, G., Furukawa, R., Sagawa, R. and Kawasaki, H.: Single-wavelength and multi-parallel dotted-and solid-lines for dense and robust active 3D reconstruction, *2019 16th International Conference on Machine Vision Applications (MVA)*, IEEE, pp. 1–6 (2019).
- [11] Ronneberger, O., Fischer, P. and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Springer, pp. 234–241 (2015).
- [12] Sadlo, F., Weyrich, T., Peikert, R. and Gross, M.: A practical structured light acquisition system for point-based geometry and texture, *Proceedings Eurographics/IEEE VGTC Symposium Point-Based Graphics, 2005.*, IEEE, pp. 89–145 (2005).
- [13] Sagawa, R., Sakashita, K., Kasuya, N., Kawasaki, H., Furukawa, R. and Yagi, Y.: Grid-based active stereo with single-colored wave pattern for dense one-shot 3D scan, *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, IEEE, pp. 363–370 (2012).
- [14] Ulusoy, A. O., Calakli, F. and Taubin, G.: One-shot scanning using de bruijn spaced grids, *Proceedings of the International Conference on Computer Vision (ICCV) Workshops*, IEEE, pp. 1786–1792 (2009).
- [15] Zhang, S. and Huang, P. S.: Novel method for structured light system calibration, *Optical Engineering*, Vol. 45, No. 8, p. 083601 (2006).
- [16] Zhang, X., Zhang, Z. and Cheng, W.: Iterative projector calibration using multi-frequency phase-shifting method, *2015 IEEE 7th international conference on cybernetics and intelligent systems (CIS) and IEEE conference on robotics, automation and mechatronics (RAM)*, IEEE, pp. 1–6 (2015).