

## 固有画像分解を用いる 奥行きと濃淡情報の複合処理

永岡 淳史 田村安孝 柳田裕隆

本研究では、Kinectを想定し、人の顔を対象とする奥行きと濃淡画像を統合処理するシステムについて、トレーニング画像の枚数を増加して検討を行った。再現性の高いトレーニング画像として、3DCGにより作成した顔の奥行き・濃淡画像を使用することとした。このトレーニング画像と、トレーニング画像に含まれていないテストデータを用い、雑音の除去と、欠落情報の復元の処理システムを構築し、性能を評価した。

### Composite processing of depth and shading information using Eigen image decomposition

Nagaoka Atsushi Yasutaka Tamura Hirotaka Yanagida

we assumed the Kinect, a system for integrated processing the depth and gray image to target a human face, was examined by increasing the number of training images. As a highly reproducible training image, it was decided to use the depth-gray image of a face that was created by 3DCG. And the training image, using test data not included in the training image, and constructs the noise removal of the restoration of the processing system of the missing information, and evaluate the performance.

## 1. はじめに

### 1-1 研究背景

近年、画像処理の分野で、関連する複数の種類のデータを複合して処理することが求められるようになってきた。その理由として、光学カメラ自体が半導体イメージセンサや周辺回路の集積化により高い性能対コスト比を持つようになったことに加え、様々なセンサも容易に利用できるようになってきたことが挙げられる。一例として、画像だけでなく奥行きやオーディオのデータを取得できる Kinect がある。Kinect はゲーム機用の入力デバイスとして開発されたため、従来は非常に高価であった三次元センサが比較的安価に利用できるようになった。このため、ゲーム以外の用途への応用も注目されている。

このようなデバイスが登場すると、様々なセンサによって取得されたデータを統合して扱う手法が求められるようになる。例えば医用画像診断の分野では、同一の患者に対し X 線、MRI、超音波など、複数の手段で取得した画像から診断情報を得る。複数のセンサにより取得されたデータは、空間分解能や精度などが互いに異なり、センサの位置もずれているため、縮尺や位置を合わせる操作が必須である。そして、複数の画像データを統合して、医師が最終的な診断情報を下す助けになるようなデータを得なければならない。

上に挙げたような、多様なデータを統合するための標準的な枠組みはまだできていない。また、医師の持つ知識という、記号として表現することが難しいものを診断システムの中に組み込む方法も、確立されていない。

図 1-1 は、本研究で考えている統合型画像処理システムの概念図である。例として人間を対象として、光学像や体表面の温度などのデータを取得する。これらのセンサとは大きく質の異なるデータの例として、人間による手描きのスケッチもシステムへの入力として考えている。(例えば、犯罪捜査などでは似顔絵が使われている。)

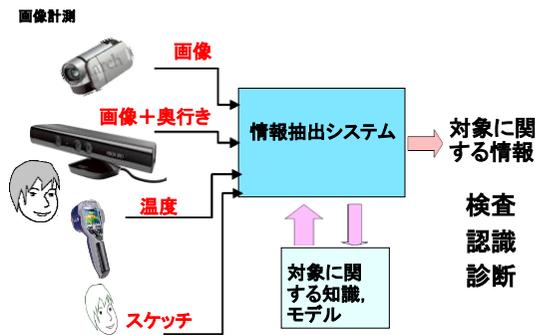


図 1-1 統合的画像処理システムの概念図

さらに、センサで取得できるデータに加え、システムは、対象である人間に関する先験的な情報である“知識”も使う。そして、最終的にシステムが出力したデータを用いて対象に関する知識を得るのである。

### 1-2 固有画像分解

この修士論文研究では、異なる複数のデータと対象に関する知識を統合する手法として固有画像分解の手法を用いる。

固有画像分解は、人の顔などの「共通する特徴」を持つデータから学習を行い、効率の高い処理を可能とする手法である。固有画像分解は顔画像の認証システムなどへの応用が検討されている。S.Chaudhuriらは、固有画像分解の手法により、人間の顔画像を対象として、雑音の除去、低解像度の画像からの高解像度の画像の復元などが可能なことを示している。

我々の研究室でも、濃淡情報と奥行き情報（永岡・卒業研究）、濃淡情報と手描きの線画（小川・修士論文研究）などの関連するデータ同士を統合して扱えることが示されている。

また、これらの研究を通して、提案する画像処理手法を検証するための Kinect によるデータ取得システムの構築を試みた。しかし、現時点では、実用に供し得るような性能評価には至っていない。これは、データ取得システムの再現性の問題があること、トレーニング画像が少ないこと、前処理システムに不備があることなどが原因となっている。

### 1-3 研究目的

Kinect を想定した、人の顔を対象とする奥行と濃淡画像を統合処理するシステムを検討する。固有画像分解を用いる統合処理の性能を評価する。雑音の除去、欠落情報の復元の性能を評価する。3DCG により作成した顔の奥行・濃淡画像を使用し、システムの再現性を高める。応用システムを作成して、動作検証を行う。

## 2. 情報復元の手法

### 2-1 学習用データ行列の作成

Kinect から得られた濃淡画像を  $g$ 、奥行き画像を  $d$  とする。 $g(m, n)$ ,  $d(m, n)$  と表される。 $(1 \leq m \leq M, 1 \leq n \leq N)$  濃淡画像と奥行き画像の画素数を  $L(MN)$  とし、図 1 のようにラスタライズを行い、ベクトル  $g, d$  に変換する。 $g, d$  を統合したベクトルを  $f$  とする。次元は  $2L$  となる。

$$f = \begin{pmatrix} g \\ d \end{pmatrix}$$

$$g = [g_1, g_2, \dots, g_L]^T$$

$$d = [d_1, d_2, \dots, d_L]^T$$

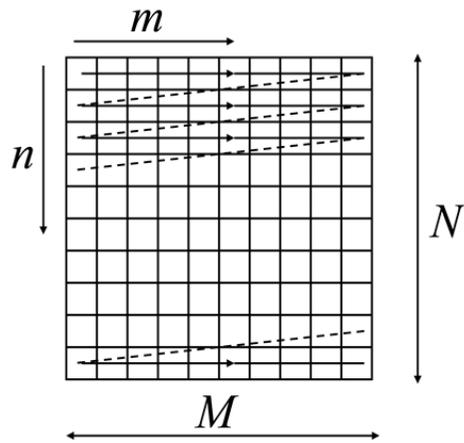


図 2-1 画像のラスタライズ

学習用の、統計的に同じ性質を持つと考えられる  $K$  個のデータ  $\{f_1, f_2, \dots, f_K\}$  を取得する。このときの  $K$  個のデータの平均画像を  $\bar{f}$  とする。

$$\bar{f} = \sum_{k=1}^K f^k$$

$\bar{f}$  を用いて学習用データ行列  $T$  を以下のように表す。

$$T = [f_1 - \bar{f}, f_2 - \bar{f}, \dots, f_K - \bar{f}]$$

この  $T$  を用いて共分散行列  $C$  をつぎのように定義する。

$$C = \frac{1}{K} T T^T = \frac{1}{K} \sum_{k=1}^K (f^k - \bar{f})(f^k - \bar{f})^T$$

## 2-2 Eigen Image の計算方法

### 2-2-1 固有値分析を使う方法

共分散行列  $C$  を固有値分析して固有値と固有ベクトル、

$$\lambda_1, \lambda_2, \dots, \lambda_K$$

$$u_1, u_2, \dots, u_K$$

を求める。

$u_1, u_2, \dots, u_K$  が固有画像 (Eigen Image) である。

しかし、共分散行列  $C$  の寸法は画素数  $\times$  画素数となり、大寸法の画像では、計算が困難になる。

### 2-2-2 特異値分解を使う方法

$2L$  行  $K$  列の学習用データ行列  $T$  は次のように特異値分解が可能である。

$$T = U \Lambda^{\frac{1}{2}} V^T$$

$$U = [u_1, u_2, \dots, u_{2L}]$$

$$V = [v_1, v_2, \dots, v_K]$$

$$A^{1/2} = \begin{bmatrix} \lambda_1^{1/2} & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \lambda_2^{1/2} & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \lambda_r^{1/2} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \end{bmatrix}$$

$A^{1/2}$ は  $r$  ( $r \leq \min(2L, K)$ ) 個の対角要素が非 0 の行列である。

一方で、トレーニング行列  $T$  から  $TT^T$  を計算すると

$$TT^T = (VA^{1/2}U^T)(UA^{1/2}V^T) = VAV^T$$

となる。

$TT^T$  の固有値分析により  $V=[v_1, v_2, \dots, v_K]$  を求めると

$$TV = UA^{1/2}V^TV = UA^{1/2}$$

となることから、 $u_1, u_2, \dots, u_K$  と  $v_1, v_2, \dots, v_K$  の間には、

$$Tv_k = \lambda_k u_k \quad k=1, 2, \dots, K$$

という関係がある。 $TT^T$  は寸法  $K/K$  であり、 $K$  はトレーニングデータの個数なので、一般に画素数  $L$  に比べ小さいため、計算は容易である。

そこで、 $Tv_k$  の計算を行い正規化することで、固有画像  $E=[u_1, u_2, \dots, u_K]$  を求めることができる。

### 3 情報復元の手法

#### 3-1 欠落した奥行画像の復元

復元を行う画像（テスト用画像と同じ統計的性質を持つ画像） $f$  から復元画像  $\hat{f}$  を求めるために、固有ベクトル  $E$  を用いて重み係数を求める。

$$W = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \end{bmatrix} = E^T (f - \bar{f})$$

上式から計算できる重み係数

$$w_n = u_n^T (f - \bar{f}) \quad (n=1, 2, \dots, K)$$

のうち  $K$  個を用いた復元画像  $\hat{f}$  は以下のように求められる。

$$\hat{f} = \sum_{k=1}^K w_k u_k + \bar{f}$$

復元の対象となる画像  $f$  として、以下のものを検討した。

#### 1) 雑音の除去

テスト画像  $f_T$  に雑音  $n$  を加えたものをシステムへの入力  $f$  として用いる。すなわち、

$$f = f_T + n$$

である。

#### 2) データの一部が欠落したもの

奥行画像あるいは濃淡画像のどちらか一方が欠落したものを  $f$  として用いる。ここで、システムへの入力は、トレーニングデータと統計的な性質が等しいと仮定している。そこで、欠落したデータの部分は、トレーニングで得られた平均画像を用いることになる。

##### a) 奥行画像が欠落したもの

システムへの入力、奥行きを表すベクトルをトレーニングで得られた $\bar{d}$ で置き換えた、

$$f = \begin{pmatrix} g_T \\ \bar{d} \end{pmatrix}$$

を使う。

b) 濃淡画像が欠落したもの

システムへの入力、濃淡画像を表すベクトルをトレーニングで得られた $\bar{g}$ で置き換えた、

$$f = \begin{pmatrix} \bar{g} \\ d_T \end{pmatrix}$$

を使う。

## 4. 評価

### 4-1 3DCGによる顔の奥行・濃淡

情報の取得<sup>[1]</sup>

最終的な応用を考えた場合、実際の人間の顔からトレーニングデータを収集する作業は欠かせない。データを収集するためには、十分な人数の、しかも統計的に同じカテゴリーに属する人を集めなければならない。また、測定の実環境や前処理システムも整備する必要がある。

しかし、このようなデータの入手は容易ではなく、環境の整備もできていない。

そこで、人体の数値モデルに基づいて、人物を含む三次元のシーンから2次元画像を生成するソフトウェアである、3DCGによる顔の奥行・濃淡情報の取得<sup>[1]</sup>が、小澤らにより

検討されている。

本研究では、3DCGにより生成した顔の濃淡および奥行データをトレーニング画像とテスト画像として使用する。

3DCGによる顔の奥行・濃淡情報の取得<sup>[1]</sup>

DAZStudio™は、DAZ 3D社製の3次元コンピュータグラフィックス(3DCG)ソフトウェアである。人体モデル(DAZ Studioではフィギュアと呼ぶ)を変形し、照明やカメラの位置を設定して、写真撮影で得られるような2次元画像や動画を作成することができる。

DAZStudioは、空気遠近法や霧などの描写用の“Atmospheric Cameras”というプラグインを追加することで、対象までの距離画像も得ることができる。これを利用して、濃淡画像と奥行データを得る。



図 4.1 DAZStudio の操作画面

図 4.1 は、DAZStudio の操作画面である。この画面では、今回のトレーニング画像とテスト画像の作成に用いた、男性キャラクターが表示されている。

図 4.2 にトレーニングに用いたデータの一例を示す。

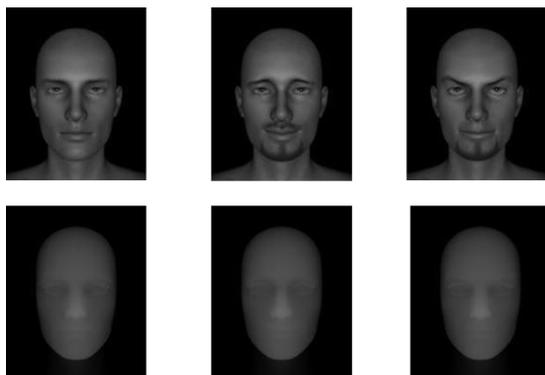


図 4.2 トレーニングデータの一例

DAZStudio の人体モデルの変形（モーフィング）の機能を利用して、この図のような顔の形状や表情を変えたデータを、100 組準備した。

#### 4-2 固有画像の計算

2-1 に示した数学的手順に基づいて、100 組の固有画像を求めた。図 4.3 に、平均画像と第一から第三までの固有画像を示す。

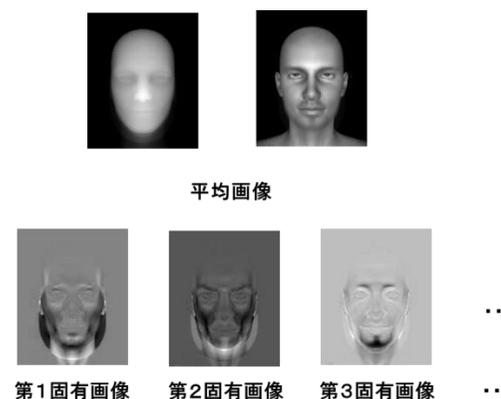


図 4.3 平均画像と固有画像

#### 4-3 画像復元システムの構築と性能評価

奥行きと濃淡画像を統合したトレーニング画像から得られた固有画像を基に、以下のような画像復元システムを構築し、その性能を評価した。

##### 4-3-1 NMSE による性能評価

復元システムの性能は NMSE (Normalized Mean Square Error, 正規化平均 2 乗誤差) により評価した。NMSE は以下の式で求められる。

$$NMSE = \frac{|\hat{f} - f_0|^2}{|f|^2} = \frac{\sum_{n=1}^{2L} (\hat{f}_n - f)_n^2}{\sum_{n=1}^{2L} f_n^2}$$

ここで、 $f_0$  は元画像、 $\hat{f}$  は復元された画像とする。NMSE は元画像  $f_0$  と復元画

像  $\hat{f}$  との差の 2 乗誤差を元の画像のエネルギーで正規化した値である。NMSE が小さければ小さいほど原画像に近い画像であり、主に画像の伝送や復元などの処理の評価に用いられる。

#### 4-3-2 画像復元

トレーニング画像から得られた固有画像を  $u_n$  ( $n = 1, 2, \dots$ ) とするとき、劣化画像

$f$  に対して、復元画像を、

$$\hat{f} = \sum_{n=1}^K u_n^T (f - \bar{f}) u_n^T + \bar{f}$$

により計算する。

##### 1) 雑音の加わった画像からの復元

3-1 に示したように、システムに入力する画像  $f$  は、テスト画像  $f_T$  に雑音  $n$  を加えたもの、

$$f = f_T + n$$

である。雑音  $n$  は、一様乱数を用いた。

まず、はじめに、劣化画像としてトレーニングに使用した画像に一様乱数を加えたものを用いて、基本的な性能の評価を行った。

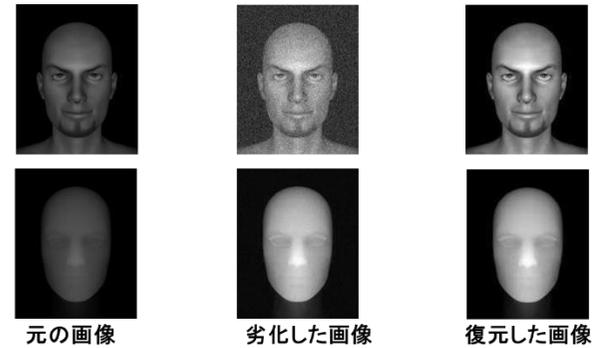


図 4-4 雑音を加えた像の復元の例

(復元した濃淡画像の NMSE :  $1.56 \times 10^{-4}$ , 使用した固有画像の個数  $K = 100$ )

図 4-4 は、復元に使用する固有画像の個数  $K$  を 100 とした場合の、元の画像、劣化画像、復元された画像の例を示す。劣化画像の NMSE は  $4.02 \times 10^{-2}$  だったのに対し、復元した濃淡画像の NMSE は、試行回数 5 回での NMSE の平均で  $1.56 \times 10^{-4}$  と改善されていることが確認できた。

図 4-5 は、復元に使用する固有画像の個数  $K$  を 10 とした場合の、元の画像、劣化画像、復元された画像の例である。劣化画像の NMSE =  $4.02 \times 10^{-2}$  に対し、復元した濃淡画像の NMSE は、 $3.80 \times 10^{-3}$  となった。

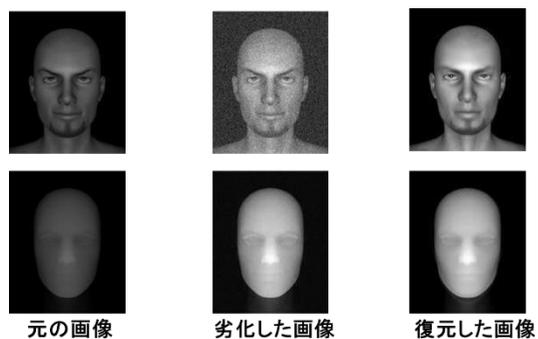


図 4-5 雑音を加えた像の復元の例

(復元した濃淡画像の NMSE :  $3.80 \times 10^{-3}$  , 使用した固有画像の個数  $K = 10$ )

図 4-6 は、復元に使用する固有画像の個数  $K$  を 1 とした場合の、元の画像、劣化画像、復元された画像の例である。劣化画像の  $NMSE = 4.02 \times 10^{-2}$  に対し、復元した濃淡画像の NMSE は、 $8.80 \times 10^{-3}$  となった。

$K=1$  では、復元された画像は、主観的にはほとんど平均画像と同じ画像となっている。

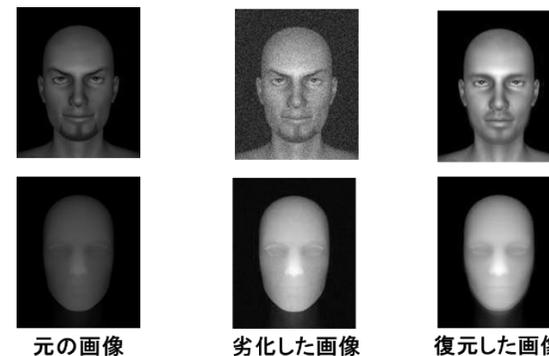


図 4-6 雑音を加えた像の復元の例

(復元した濃淡画像の NMSE :  $8.80 \times 10^{-3}$  , 使用した固有画像の個数  $K = 1$ )

#### 4-3 固有画像の個数と NMSE の関係

復元された画像の濃淡画像に関する NMSE 値と、復元に用いた固有画像の個数との関係をグラフに示したものが図 4-7 である。固有画像の枚数を増加していくと、NMSE 値は急速に減少し、その後、減少の幅は小さくなっていくことが見て取れる。復元した濃淡画像の NMSE は  $1.56 \times 10^{-4}$  ( $K=100$ )  $\sim 8.80 \times 10^{-3}$  ( $K=1$ ) となっている。

このグラフより、今回使用したトレーニングデータに関しては、固有画像の個数は、90 あれば十分であることが確認できる。

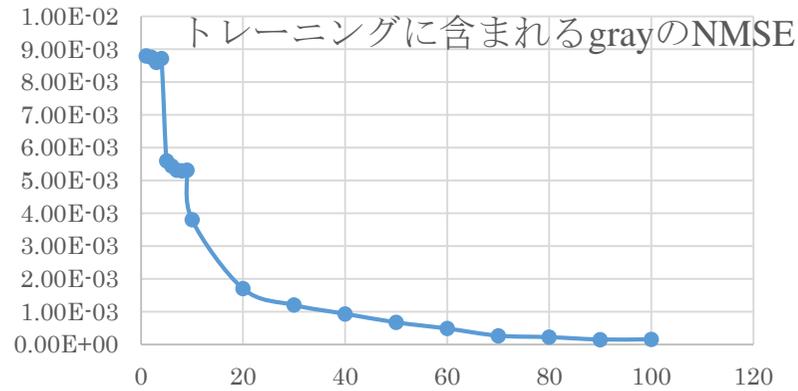


図 4-7 復元された濃淡画像の NMSE と固有画像の個数の関係

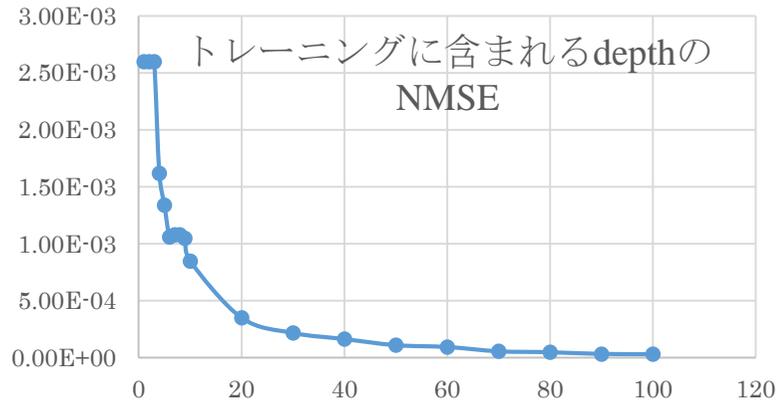


図 4-8 復元された奥行き画像の NMSE と固有画像の個数の関係

次に、劣化画像として、トレーニングに使用していない画像に一樣乱数を加えたものを使って、画像を復元し、性能を評価した。

図 4-9 は、復元に使用する固有画像の個数  $K$  を 100 とした場合の、元の画像、劣化画像、復元された画像の例である。劣化画像の NMSE は  $4.02 \times 10^{-2}$  だったのに対し、復元した濃淡画像の NMSE は、試行回数 5 回での NMSE の平均で  $1.34 \times 10^{-3}$  と改善されている。しかし、劣化画像をトレーニングに使用した画像から生成した場合の復元画像の評価値  $1.56 \times 10^{-4}$  に比べると、NMSE 値の改善率は低いことが確認できた。

復元に使用する固有画像の個数  $K$  を 10, 1 とした場合の、元の画像、劣化画像、復元された画像の例を図 4-10, 図 4-11 に示す。

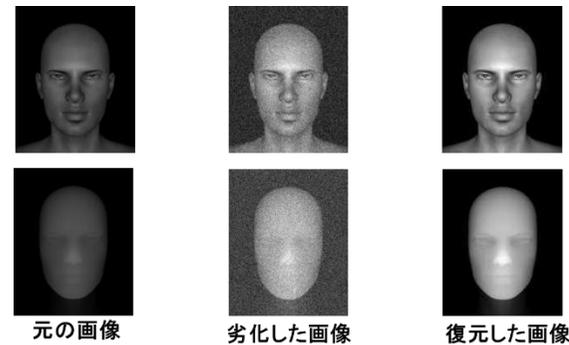


図 4-9 雑音を加えた像の復元の例

(復元した濃淡画像の NMSE :  $1.34 \times 10^{-3}$ , 使用した固有画像の個数  $K = 100$ )

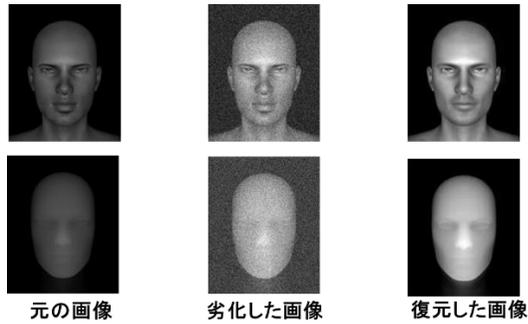


図 4-10 雑音を加えた像の復元の例

(復元した濃淡画像の NMSE :  $1.34 \times 10^{-3}$  , 使用した固有画像の個数  $K = 10$ )

復元した濃淡画像の NMSE :  $6.34 \times 10^{-3}$

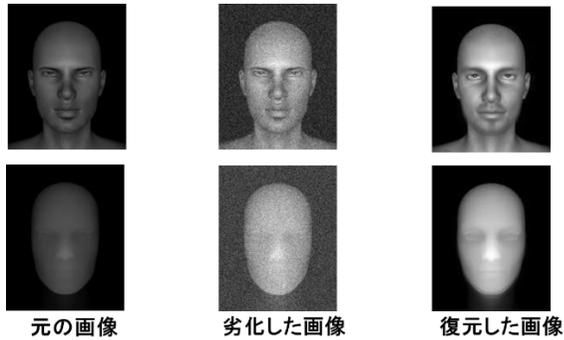


図 4-11 雑音を加えた像の復元の例

(復元した濃淡画像の NMSE :  $1.44 \times 10^{-2}$  , 使用した固有画像の個数  $K = 1$ )

固有画像の個数と NMSE の関係

復元された画像の濃淡画像に関する NMSE 値と、復元に用いた固有画像の個数との関係をグラフに示したものが図 4-12, 図 4-13 である。固有画像の枚数を増加していくと、

NMSE 値は急速に減少し、その後、減少の幅は小さくなっていく傾向は、劣化画像をトレーニングに使用したデータの中から選んだ場合と同じである。しかし、復元した濃淡画像の NMSE は  $1.34 \times 10^{-3}$  ( $K=100$ )  $\sim$   $1.44 \times 10^{-2}$  ( $K=1$ ) , 奥行き画像の NMSE は  $2.24 \times 10^{-4}$  ( $K=100$ )  $\sim$   $2.82 \times 10^{-3}$  ( $K=1$ ) であり、NMSE 値は大きくなっている。

また、奥行き画像に関して、NMSE 値と復元に使用した固有画像の個数が単調減少とならず、固有画像の個数を増やすと NMSE が増加する点がある。

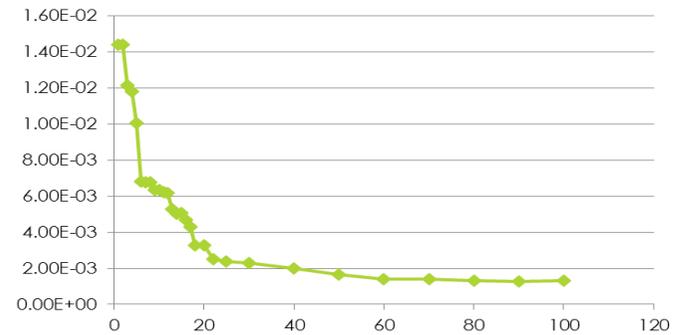


図 4-12 復元された濃淡画像の NMSE と固有画像の個数の関係

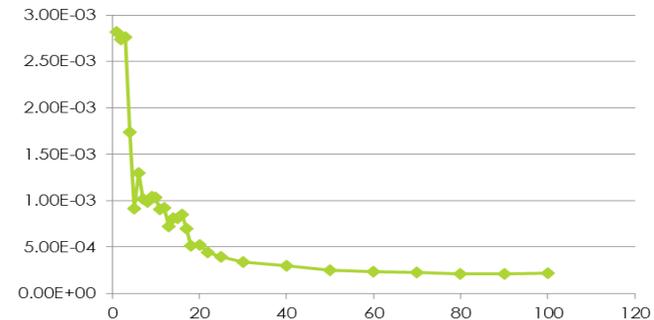


図 4-13 復元された奥行き画像の NMSE と固有画像の個数の関係

## 雑音除去についてのまとめ

雑音除去の効果を NMSE により評価した。トレーニングに使用した画像では、

濃淡画像 :  $4.02 \times 10^{-2} \rightarrow 1.31 \times 10^{-4}$

奥行画像 :  $6.10 \times 10^{-2} \rightarrow 1.74 \times 10^{-5}$

という結果となり NMSE は 2 桁程度減少した。主観的にも大きな雑音除去効果が認められる。

一方、トレーニングに使用しなかった画像では、

濃淡画像 :  $3.91 \times 10^{-2} \rightarrow 1.30 \times 10^{-3}$

奥行画像 :  $6.11 \times 10^{-2} \rightarrow 1.90 \times 10^{-4}$

となり、トレーニングデータ内の画像を劣化画像として選んだ場合に比べ 1 桁程度 NMSE 値は増えるものの、NMSE を減少できることが確認された。

## 2) 欠落した情報の復元

奥行または濃淡のどちらかのデータが欠落している場合に、それを復元する処理システムを構築し、性能を評価した。

### a) 奥行画像の復元

システムへの入力、奥行きを表すベクトルをトレーニングで得られた  $\bar{d}$  で置き換えた、

$$f = \begin{pmatrix} \mathbf{g}_T \\ \bar{d} \end{pmatrix}$$

を使う。

### b) 濃淡画像の復元

システムへの入力、濃淡画像を表すベクトルをトレーニングで得られた  $\bar{g}$  で置き換えた、

$$f = \begin{pmatrix} \bar{g} \\ \mathbf{d}_T \end{pmatrix}$$

を使う。

図 4-14 は、復元に使用する固有画像の個数  $K$  を 100 とした場合の、元の画像、劣化画像、復元された画像の例を示す。復元された奥行き画像の、顔の正中線に相当する直線上のプロファイルのグラフも示す。劣化画像の NMSE は  $4.02 \times 10^{-2}$  だったのに対し、復元した奥行き画像の NMSE は、 $1.2 \times 10^{-3}$  と、ほぼ 1/4 になった。

図 4-15 は、復元に使用する固有画像の個数  $K$  を 10 とした場合の、元の画像、劣化画像、復元された画像の例である。劣化画像の NMSE= $4.02 \times 10^{-2}$  に対し、復元した濃淡画像の NMSE は、 $1.4 \times 10^{-3}$  となった。

図 4-16 は、復元に使用する固有画像の個数  $K$  を 1 とした場合の、元の画像、劣化画像、復元された画像の例である。劣化画像の NMSE= $4.02 \times 10^{-2}$  に対し、復元した濃淡画像の NMSE は、 $2.2 \times 10^{-3}$  となった。

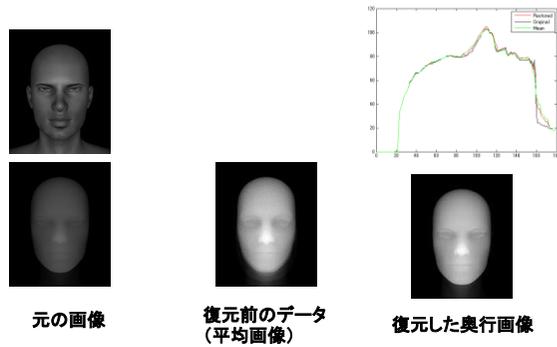


図 4-14 欠落した情報（奥行）の復元の例

(復元した奥行画像の NMSE :  $3.80 \times 10^{-3}$  , 使用した固有画像の個数  $K = 100$ )

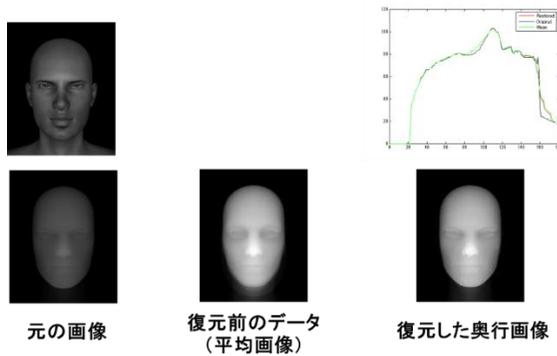


図 4-15 欠落した情報（奥行）の復元の例

(復元した奥行画像の NMSE :  $3.80 \times 10^{-3}$  , 使用した固有画像の個数  $K = 10$ )

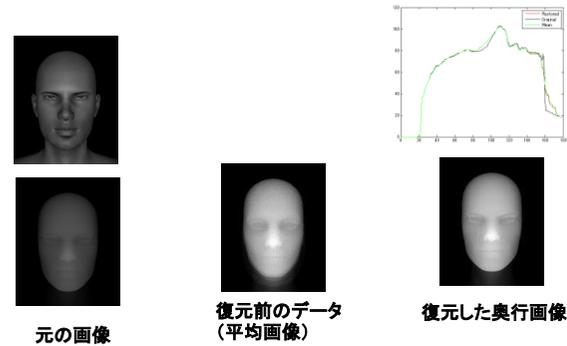


図 4-16 欠落した情報（奥行）の復元の例

(復元した奥行画像の NMSE :  $3.80 \times 10^{-3}$  , 使用した固有画像の個数  $K = 1$ )

#### 固有画像の個数と NMSE の関係

復元された奥行画像に関する NMSE 値と、復元に用いた固有画像の個数との関係をグラフに示したものが図 4-17 である。固有画像の枚数を増加していくと、いったん減少した NMSE 値が増加し、固有画像の個数が 30 以上で、ほぼ一定値になる。また、NMSE の低下も、復元した奥行画像の NMSE は  $1.34 \times 10^{-3}$  ( $K=100$ )  $\sim$   $1.44 \times 10^{-2}$  ( $K=1$ ) , 奥行き画像の NMSE は  $2.24 \times 10^{-4}$  ( $K=100$ )  $\sim$   $2.82 \times 10^{-3}$  ( $K=1$ ) であり、NMSE 値は大きくなっている。



図 4-17 復元された奥行画像の NMSE と固有画像の個数の関係

#### 欠落情報の復元についてのまとめ

奥行画像の欠落したデータからの復元の効果を NMSE で評価した。トレーニングに使用しなかった画像では、奥行画像の NMSE は  $2.2 \times 10^{-3}$  から  $1.2 \times 10^{-3}$  と小さくなった。しかし、雑音の除去に比べると、NMSE を小さくする効果は非常に小さいと言える。

また、固有画像の個数に対する NMSE の変化も、雑音を加えた画像の復元に比べ、30 程度と、少ない個数で飽和していることがわかる。

### 5. 応用システムの構築

#### Kinect を用いた任意視点の画像生成システム

この章では、濃淡と奥行の 2 つの画像データを取得できる Kinect を用いた応用システムの構築について述べる。構築したシステムは、濃淡画像と奥行画像を統合し、任意視点からの 2 次元画像を生成する。

このようなシステムの応用としては、以下のような事例を考えることができる。

- ・視点位置の補正

Kinect を用いて取得した人物の濃淡・奥行データから、視点の位置を補正した濃淡画像を得る。これにより、画像認識の精度が向上できる。

- ・異なる視点からの画像

データ取得時とは異なる視点からの人物像を生成できる。これは、映像の記録、報道やエンターテインメントなどの分野で有用と考えられる。本研究では、顔の濃淡と奥行のデータに限定して検討をしているが、一般的な建造物などにも応用できれば、一度取得したデータから任意の視点から撮影した画像を再現することができる。

#### 5-1 Kinect センサの概要

図 5-1 は Kinect センサ (ver.1) の外観である。図で Color センサとあるのは、通常のカラーカメラで濃淡画像の取得に用いる。IR プロジェクタと IR センサは空間的にランダムなパターンを赤外線により対象に投影し、それを検出することで対象表面の奥行を計算する。

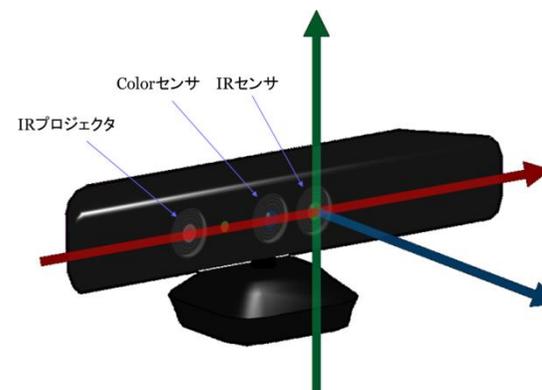


図 5-1 Kinect センサの外観 (ver.1)

図 5-2 は IR センサ、カラーセンサの位置と座標系の配置を示す図である。一方、図

5-3 は、2つのセンサの焦点位置と取得された画像データとの関係を示す投影モデルである。各センサは光学レンズによる撮像系を構成するが、撮像系のモデルとしては、座標系内に設定された焦点に、対象表面からの光線が集まり、これが焦点と対象の間に設定される仮想的な焦点面に投影されると考える。

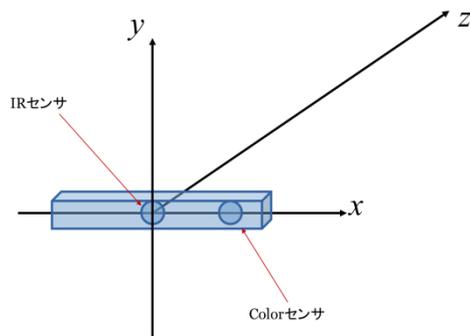


図 5-2 Kinect センサの 2つのセンサ

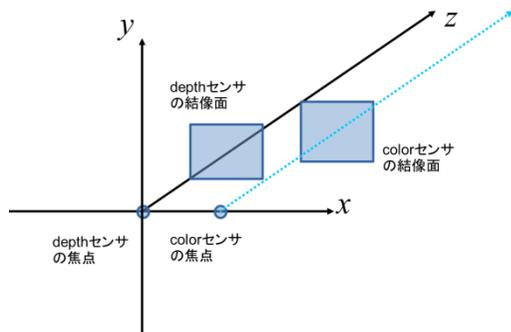


図 5-3 焦点と結像面

以下で、任意視点の画像を生成する手順について説明する。

1) 奥行が計測できた点群の 3D 座標を計算

Kinect センサにより、奥行が計測された複数の点 (point cloud) が得られる。point cloud 中の 1 点は、対応する奥行センサの結像面上の 1 点と、奥行センサの焦点を結ぶ直線上に存在する。この直線がカラーセンサの結像面と交差する点の色情報が、point cloud 上の点に対応づけられる。

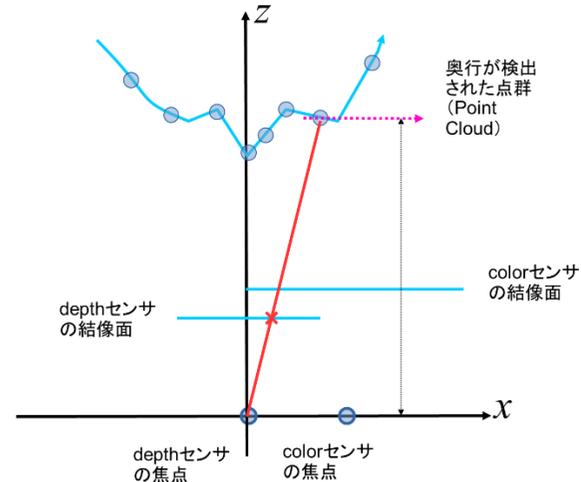


図 5-3 奥行が計測された点群と結像面の関係

3) 点群と色情報を結合

1) の操作を point cloud に属する全ての点について行うことで、これらの点の色情報と奥行データが結合される。

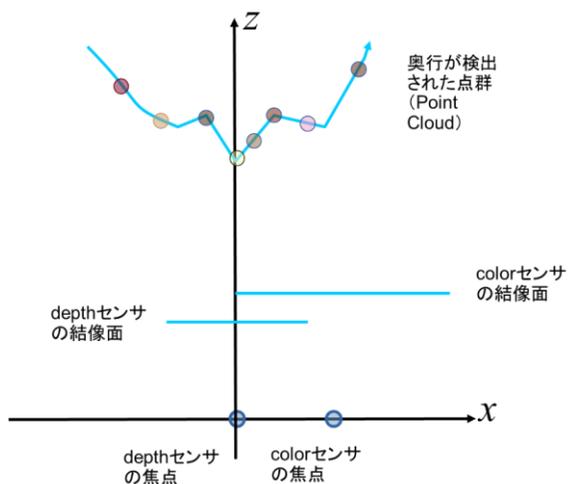


図 5-4 点群の全ての点の奥行と色情報が計測された状態

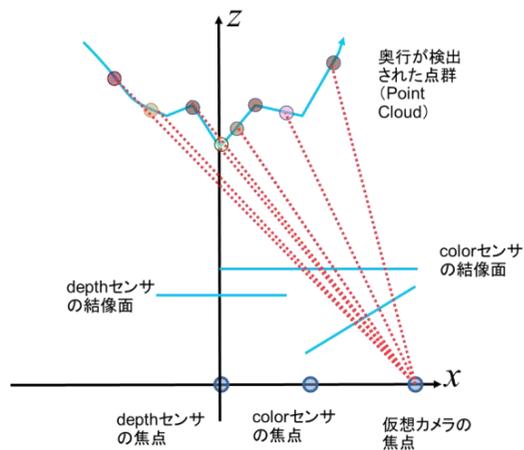


図 5-5 仮想カメラの結像面に点群の色情報を投影

3) 仮想カメラから見た像を生成

奥行と色情報が結合された点群の色情報を仮想カメラの結像面に投影することで、任意視点からの画像が生成される。

Kinect による動作検証

Kinect を用いて人物のカラー画像と奥行画像を取得し、上述した手順にしたがって、任意視点の画像を得るシステムを構築した。ソフトウェアの開発は MATLAB を用いた。



カラー画像

奥行画像

図 5-6 Kinect で得られたカラー画像と奥行画像

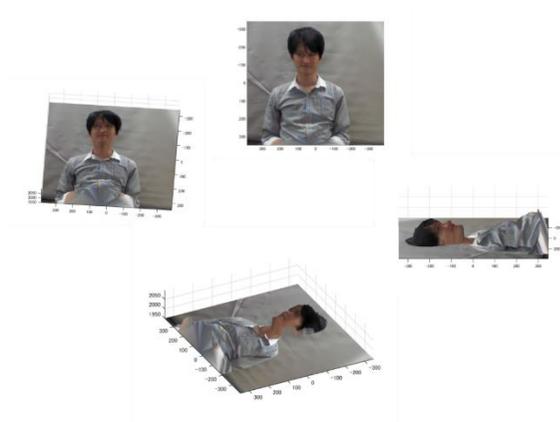


図 5-7 任意視点からの画像の例

図 5-6 は、Kinect を用いて得られた人物のカラー画像と奥行画像である。これを基に生成した任意視点からのカラー画像を図 5-7 に示す。

奥行の計測が可能であった点については、ほぼ、異なる視点からの画像が生成できる。しかし、人物画像の一部（例えば鼻の部分）は、視点を変えると、立体的形状が明らかに実際のものとはずれていることがわかる。これは、奥行計測の誤差のためと考えられる。

また、当然ではあるが、奥行が計測できなかった部分は、カラー情報が欠けている。

このような不完全な計測システムに対し、4 章までに述べたデータの復元システムを用いることで、より実際に近い、任意視点からの画像が生成できると考えている。

## 6.まとめ

Kinect を想定し、人の顔を対象とする奥行と濃淡画像を統合処理するシステムについて、トレーニング画像の枚数を増加して検討を行った。再現性の高いトレーニング画像として、3DCG により作成した顔の奥行・濃淡画像を使用することとした。3DCG ソフト

ウェアとしては DAZStudio™ を用いた。奥行画像は空気遠近法や霧などの描写用の “Atmospheric Cameras” を使用した。トレーニング画像として濃淡画像と奥行画像のペアを 100 組作成した。

このトレーニング画像と、トレーニング画像に含まれていないテストデータを用い、雑音の除去と、欠落情報の復元の処理システムを構築し、性能を評価した。トレーニングでは、100 組のデータから、奥行きと濃淡画像を統合した固有画像を抽出し、それを基に、雑音の加わった画像からの復元と欠落した情報の復元を行い、性能を NMSE (Normalized Mean Square Error) により評価した。

その結果、雑音の除去では、NMSE が  $3.91 \times 10^{-2}$  から  $1.30 \times 10^{-3}$  まで改善された。欠落情報の復元では、濃淡画像または奥行画像の一方を欠落させたデータを入力し、元のデータの復元を試みた。NMSE は、濃淡画像のみからの奥行画像の復元では  $2.2 \times 10^{-3}$  から  $1.2 \times 10^{-3}$  まで、奥行画像のみから濃淡画像の復元では、 $1.44 \times 10^{-2}$  から  $1.29 \times 10^{-2}$  と改善された。

また、濃淡と奥行の両方を利用する応用システムとして、Kinect を用いる任意視点からの画像を生成するシステムを構築した。実際の人物から取得したデータにより、任意視点からの画像の生成が可能であることを示した。

この応用システムに、4 章までで提案したデータ復元手法を用いることで、より妥当な画像が生成されると期待している。しかし、4 章までの成果を組み込んだ応用システムは未完成であり、動作検証は行われていない。

## 7.謝辞

本研究を進めるにあたって、田村安孝・柳田裕隆教授には大変お世話になりました。始めは研究をどのように進めればよいかわかりませんでしたが、田村・柳田教授の研究の方針、プログラム作成などさまざまな分野で助言や指導のおかげで研究を成し遂げるこ

とことができました。改めて感謝を申し上げます。

#### 8.参考文献

- [1]小川達也,Eigen Image Decomposition に基づく複合的画像処理システム,(2014)
- [2]Subhasis Chaudhuri and Joshi Manjunath: Motion-free super-resolution, pp148-172,SPringer (2005)
- [3]Matthe Turk and Alex Pentland: Eigenfaces for Recognition: Journal of Cognitive Neuroscience Volume3, Number 1(1991)
- [4]Xiaou Tang and Xiaogang Wang: Face Sketch Synthesis and Recognition, (2003)
- [5]Jun Liu,Song Chen: Fractional Order singular value decomposition representation for face recognition(2008)