

視点依存形状最適化による高精細自由視点画像生成

高井勇志[†] 延原章平[†] 田辺太一[†] 松山隆司[†] ウ小軍^{††}

[†] 京都大学

京都市左京区吉田本町

^{††} 日本電信電話株式会社 NTT サイバースペース研究所

神奈川県横須賀市光の丘 1-1

E-mail: [†]{takai,nob,tanabe,tm}@vision.kuee.kyoto-u.ac.jp, ^{††}wu.xiaojun@lab.ntt.co.jp

あらまし 3次元ビデオ技術を用いた新たな自由視点画像生成法として、画像生成時の仮想カメラ視点に応じた局所的形状最適化と、高ズームカメラの撮影画像を優先的に利用することによる高精細テクスチャ生成法を組み合わせたアルゴリズムを提案する。前者の局所的形状最適化によって、カメラキャリブレーションや形状復元の精度のため大局的最適化では頭打ちとなる photo-consistency の限界を超えた形状の最適化が行え、映像化時のテクスチャのずれやボケを低減させることが可能となる。また、後者のテクスチャ生成法によって、高解像度の撮影画像をテクスチャ生成に有効に用いることが可能となる。つまり、多少の誤差が含まれた入力データであっても、局所的な最適化を施すことによって高精度の形状と高精細なテクスチャの両方を同時に得ることができる手法である。本論文では、実画像を用いて定性的・定量的評価を行い、既存の手法との比較や考察によって提案手法の有用性を示す。

キーワード 3次元ビデオ, 高精細自由視点画像生成, 多視点画像, 視点依存形状最適化, 大局的最適化, 局所的最適化

1. はじめに

3次元ビデオ技術を用いて生成した自由視点画像の品質は、大きく、(1)入力多視点画像の有効解像度^(注1)、(2)カメラキャリブレーションの精度、(3)形状復元の精度、の3つの要素によって決定される。

入力多視点画像の有効解像度については、カメラのセンサー解像度が一定である場合、カメラのズーム値(焦点距離)によって決定される。ただしズームを行った場合には、カメラの視野角および被写界深度が狭まるため、必然的に撮影可能範囲が減少する。したがって、

- カメラ台数を増やす。
- 能動カメラによる追跡撮影を行う[1]。

などの方法によって撮影可能範囲を補償することが必要となる。本研究では、全身撮影を行う標準ズームカメラ複数台と顔等の注目部位を撮影する高ズームカメラを組み合わせることによって、十分な撮影可能範囲と有効解像度を得ているとして、この問題については議論の対象外とする[2]。

本研究が対象とするのは、カメラキャリブレーションと形状復元の精度の問題を踏まえた高精細な自由視点画像の生成法である。以下では、まず始めに3次元ビデオ技術を用いた自由視点画像生成処理についてその概要を記し、上記の問題点について述べていく。

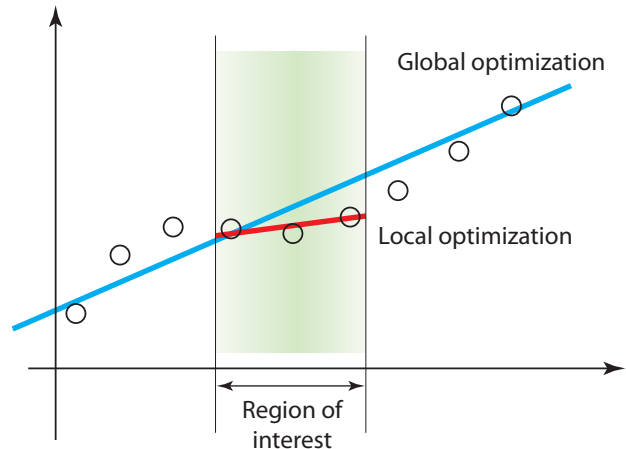


図1 大局的・局所的最適化の概念図。○は観測データを表しており、この分布に大局的・局所的に直線(モデル)当てはめを行う課題の例を示している。詳細は本文を参照。

(1) 撮影スタジオに複数のカメラを設置し、適切な焦点距離(ズーム値)を設定する。

(2) 各カメラの内部・外部パラメータを求めるが、現実問題として完全に正確なカメラパラメータを得ることは容易ではないため、求めたパラメータには誤差が含まれている。

(3) 撮影画像とカメラパラメータを元にして3次元形状復元を行う。形状復元法には、テクスチャマッチングに基づくもの[3]~[5]、シルエットに基づくもの[6]~[9]、これらを組み合わせたもの[10]~[14]、などいくつ

(注1): 撮影画像を対象の3次元形状上に投影した時の解像度。

かの方法があるが、いずれの手法であっても完全に正確な形状を得ることは困難である。なぜならば、手法そのものの復元精度の問題に加えて、カメラパラメータの正確さも形状復元に大きな影響を与えるためである。したがって、得られた形状はカメラパラメータの誤差と復元精度を含有した状態で大局的に最適化されたものとなっている。

(4) 最後に、撮影画像から生成したテクスチャを得られた3次元形状に貼付け、任意の視点から見た対象の見えを生成する。ここで単純な方法を用いると、上記の全ての誤差や復元精度の問題が、テクスチャのずれやボケといった見えの悪さとして、まとめて表出することになる。

このカメラキャリブレーションや形状復元の精度の問題によって引き起こされる見えの悪さを改善する一手法として、画像生成時の仮想カメラ視点に応じて見えを修正する映像化、いわゆる視点依存テクスチャ最適化法が提案されている[15],[16]。この方法の基本的な考え方は以下の通りである。

- カメラパラメータや形状には誤差が存在するが、これらを改善しようとするのではなく、仮想カメラ視点に応じて撮影画像を変形させることによって、生成するテクスチャのずれやボケを少なくする。

- 仮想カメラに近い^(注2)実カメラが撮影した画像を優先的に用いてテクスチャを生成することによって、撮影時の対象の見えを可能な限り再現する。

この視点依存テクスチャ最適化法は、最終的な出力画像の見えを評価することによる直接的な方法であるため、自然にずれやボケを低減させることが可能である。また、仮想カメラの移動に応じてそれに近い実カメラの撮影画像を優先的に用いるため、視点に応じて対象表面の見えが変化する鏡面反射効果を再現できるという利点がある。

しかし、仮想カメラを移動させながら複数フレームの映像化を行ったときに、テクスチャが対象表面を不自然に滑るように見える現象が生じることがある。これは、対象形状を考慮せずに画像変形のみでテクスチャの辻褄を合わせているため、真の形状と復元形状との差がある箇所においてテクスチャを一致させるように変形させた結果、テクスチャが部分的に対象表面上を移動してしまうためだと考えられる。

上記に対して、本研究では以下のような考えに基づいてアルゴリズムを設計する。

(1) 通常形状復元は“どの方向から見ても誤差が一樣に少なくなる(photo-consistentである[4])こと”を目的関数とした、大局的最適化による解である。これに対して、“仮想カメラから見える範囲のみに関して誤差が少なくなること”に特化した最適化、つまり局所的最適

化を行えば、その範囲に対して、よりphoto-consistentな形状を復元できることが期待される。ここで、図1に大局的/局所的最適化の概念図を示す。この図から明らかのように、大局的最適化を行うと全体として残差が最も少なくなるようにモデル当てはめが行われているが、注目範囲を定めて局所的に最適化を行うと、モデルを変更すること無しに、より高い精度でモデル当てはめができることが分かる。この図により、視点に依存した局所的な形状最適化を行う提案手法の論理的有効性が示されているといえる。

(2) 仮想カメラとの視線方向の近さのみではなく、対象表面上におけるテクスチャの有効解像度が高くなるような撮影画像を持つ実カメラを優先的に選択することによって、高ズームカメラの画像を有効的に用いることができる。

本研究では、前者を視点依存形状最適化、後者を有効解像度を考慮した視点依存テクスチャマッピングとして、これらを組み合わせることによって、高精度な形状復元と高精細なテクスチャマッピングを実現する。

以下、本稿では、提案アルゴリズムの詳細について述べた後、実画像を用いた定性的・定量的評価を行い、最後に考察と議論を行う。

2. 提案アルゴリズム

2.1 概要

前節にて述べたように、本研究の入力データは、キャリブレーション済みの複数の(異なる焦点距離を持った)実カメラ $c_i \in C$ ($i = 1, \dots, n$)、これらによって撮影された多視点画像 I_{c_i} 、復元された3次元形状(大局的最適化形状) S 、仮想カメラ \hat{c} 、である。本提案手法は、これらのデータを用いて、まず視点依存形状最適化を行い、次に有効解像度を考慮した視点依存テクスチャマッピングを行うことによって、高精細な自由視点画像を生成する。以下の節では、それぞれの手法の詳細について述べていく。

2.2 視点依存形状最適化

本研究における視点依存形状最適化の処理の概要は以下の通りである(各記号と仮想カメラ、対象形状の関係については図2を参照)。

Step 1. S を \hat{c} に投影して、各画素 p の奥行き値 d_p を計算する。また、 p に投影された S 上の点を q 、その世界座標系での位置を q_w とする。

Step 2. \hat{c} の投影中心の世界座標系における位置を $O_{\hat{c}}$ 、 \hat{c} の画像(つまり出力画像)の画素 p の世界座標系における位置を p_w としたとき、 $O_{\hat{c}}$ を起点として p_w へ向かう直線(視線)を $l_p = t e_p + O_{\hat{c}}$ とする。ただし、 t は媒介変数、 e_p は視線方向の単位ベクトルである。このとき、 q の世界座標系での位置は $q_w = d_p e_p + O_{\hat{c}}$ となる。

(注2): 視線方向の内積によって定義される距離尺度による。詳しくは後述する。

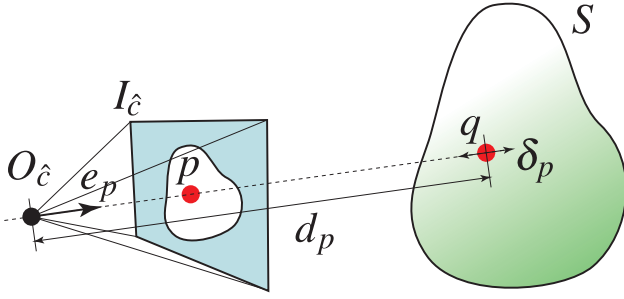


図2 仮想カメラと対象形状の幾何学的関係

Step 3. 画素 p に対応する S の奥行き値 d_p を視線 l_p に沿って距離 δ_p だけ変更させた場合、つまり、 q の世界座標系での位置を $(d_p + \delta_p)e_p + O_{\hat{c}}$ としたとき、その変更量の良さを photo-consistency に基づく関数 $E_p(\delta_p)$ によって評価する。

Step 4. $E_p(\delta_p)$ を最小化するように奥行き値 d_p の変更量 δ_p を最適化することによって、 S を \hat{c} から見た時に最適となる形状へと変形する。

この最適化された形状が、通常の3次元復元法によって得られた大局的最適化形状に対する、局所的最適化形状である。

上記の処理の際に考慮すべき点は以下の2点である。

(1) photo-consistency に基づく関数 $E_p(\delta_p)$ の設計。

(2) $E_p(\delta_p)$ を画素毎に最適化するのではなく、3次元形状の連続性を保障すること。

以下では、これらの点について述べていく。

2.2.1 Photo-consistency 関数

一般に photo-consistency を計算するためには、評価対象となる空間中の微小平面の位置・大きさ・法線方向と、その微小平面をどの実カメラが観測可能かという情報が必要となる。今回は

- 位置: $(d_p + \delta_p)e_p + O_{\hat{c}}$
- 法線: $\delta_p = 0$ のときの q の法線方向 n_p にて近似
- 大きさ: 人手により固定値 κ を設定
- 観測可能カメラ: $\delta_p = 0$ のときの q を観測可能なカメラの集合 $C_q \subset C$ にて近似

とする。ただし C_q は、 S を用いた state-based-visibility [5], [12] によって求め、奥行きの変更量 δ_p に対して不変であることに注意する。これは、特に $\delta_p > 0$ のときに点 q が S の内側へと移動することから、その位置で state-based-visibility を計算するとどのカメラからも観測されず $C_q = \emptyset$ となる可能性が高いためである。

δ_p に対応する上記の微小平面 π_{δ_p} を用いて、photo-consistency 関数 $E_p(\delta_p)$ を下記のように定義する。

$$E_p(\delta_p) = 1 - \exp\left[-\tan^2\left(\frac{\pi}{4}(\rho(\delta_p) - 1)\right)/\sigma^2\right] \quad (1)$$

ここで、 $\rho(\delta_p) = \frac{1}{|C_q|C_2} \sum_{c_i, c_j \in C_q} \text{ZNCC}(\pi_{\delta_p}, c_i, c_j)$ であり、 $\text{ZNCC}(\pi_{\delta_p}, c_i, c_j)$ は、微小平面 π_{δ_p} をカメラ c_i と c_j の画

像上に投影した位置での画素値間の ZNCC (zero-mean normalized cross correlation) である。また、 $|C_q|C_2$ はカメラ集合 C_q から2台のカメラを選ぶ組み合わせの数であり、 σ は元となった ZNCC の値のどの範囲をより photo-consistency に反映させるかを調節するパラメータである。例えば $\sigma = 0.5$ のとき $[-1, 0]$ の範囲の値が概ね 1.0、つまり最悪値へと変換されるが、 $\sigma = 2$ のときは $[-1, -0.75]$ の範囲の値が概ね 1.0 へと変換される。後述の実験ではこの値は人手によって決定した。

2.2.2 最適化時における形状の連続性の保障

$E_p(\delta_p)$ を画素毎に最適化した場合、“より photo-consistency の高い位置”へと各画素の奥行き値が独立に変更されるが、このことは必ずしも映像化の品質向上に寄与するとは限らない。なぜならば、奥行き値を探索した結果、“真値とは全く別の奥行き値にて偶然 photo-consistency が高くなってしまふ”場合、周囲の奥行き値との連続性がない棘皮状の形状になってしまう。

このような問題を解決するために、本研究では下記の2点の制約項を導入する。

- 奥行き値変更量の絶対値に対するペナルティ: $E_d(\delta_p) = |\delta_p|$.
- 隣接^(注3)する画素 p, p' 間における奥行き値の変更量の連続性制約: $E_c(\delta_p, \delta_{p'}) = |\delta_p - \delta_{p'}|$.

前者は、元となる大局的最適化形状は充分真の形状に近いことからより変更量の少ない形状を優先することを意味し、後者は、対象表面形状は滑らかに変化することを意味している。つまり、最適化の結果において表面の滑らかさを保つような制約となっている。

これらと前述の $E_p(\delta_p)$ を組み合わせ、最適化の目的関数を下記のように定義する。

$$E = \lambda_p \sum_{p \in P} E_p(\delta_p) + \lambda_d \sum_{p \in P} E_d(\delta_p) + \lambda_c \sum_{p, p' \in N} E_c(\delta_p, \delta_{p'}) \quad (2)$$

ただし、 P は全画素の集合、 N は画素間の連結関係の集合であり、 $\lambda_p, \lambda_d, \lambda_c$ は各項の重み係数である。

この目的関数 E を最小化するような各画素の奥行き値の変化量 δ_p を求めることによって、より photo-consistency が高く、かつ、妥当な形状への最適化が達成される。

しかしながら、一般にこの形式の関数を最小化することは困難な問題である。そこで本研究では、 δ_p が離散かつ限られた値を取るものとして、離散最適化問題(組み合わせ最適化問題)へと変換し、belief-propagation によって解を得る。こうして大局的最適化形状 S の仮想カメラ \hat{c} における深度マップ $\hat{D} = \{d_p, p \in P\}$ の局所最適化を行うと、深度マップ \hat{D}' が得られる。これが、求められた局所的最適化形状を表している。ただし、画素 p, p' 間の連結関係 N には巡回構造が存在するため、

(注3): 本研究では隣接は4連結で考える。

loopy-belief-propagation [17] となり厳密な大域最適解を得ることはできない。しかし、この点については実用上問題ないことを実験により確認した。

2.3 有効解像度を考慮した視点依存テクスチャマッピング

前節のアルゴリズムによって、仮想カメラ \hat{c} における局所的最適化深度マップ \hat{D} を得た。本節の目的は、この \hat{D} にテクスチャを与えて仮想カメラ \hat{c} から見た対象を映像化することである。

ここで重要となるのは“どのような基準で \hat{D} にテクスチャを与えるか”という点である。すなわち、 \hat{D} 上の各点 q は複数の実カメラによって撮影されているはずであるが、それらの多視点画像からどのように q の色を生成するかを考える必要がある。

この問題に対して、従来の視線依存レンダリング手法は、仮想カメラ \hat{c} の視線方向により近い視線を持つ実カメラによって撮影された画像を優先して用いることによって、テクスチャの透視投影による変化量をより小さくし、結果的に撮影画像に近い品質の映像化を行うことを目指したものである。

しかしこの基準では、撮影画像の有効解像度が考慮されていないため、仮に高ズームの実カメラと通常ズームの実カメラが仮想カメラに対して同じような距離にあったとしても、通常ズームカメラに仮想カメラが少しでも近い場合はその撮影画像が用いられてしまうために、高ズームカメラによって撮影された画像を有効に活用することができない。

この問題を解決するため、本研究では、視線の近さと対象表面での有効解像度の両者を同時に扱う尺度 $g_p(c)$ を以下のように定義する。

まず、画素 p における奥行き値が d_p として得られており、この奥行きを持つ点 q を実カメラ $c \in C_q$ が観測可能であるとする。実カメラ c の光軸方向を z_c 、仮想カメラ \hat{c} の光軸方向を $z_{\hat{c}}$ とすると、その内積 $z_c \cdot z_{\hat{c}}$ を基準として“視線方向の一致度”を定義する。これが従来の視点依存レンダリングにおけるカメラ選択尺度である。

次に、点 q における実カメラ c の画像の有効解像度 r_c を以下の手順で計算する（各記号と仮想カメラ、実カメラ、対象形状の関係については図3を参照）。

(1) 画素 p に隣接する画素 p' とし、 p' に対応する対象表面上の点を q' とする。

(2) 点 q および点 q' を実カメラ c に投影した画像上の点を、それぞれ p_c, p'_c とする。

(3) p_c と p'_c との間の距離を $|p_c - p'_c|$ とする。画素 p に隣接する全ての画素についてこの距離を計算し、最も小さい距離を点 p における実カメラ c の画像の有効解像度 r_c と定義する。

以上の視線方向の一致度 $z_c \cdot z_{\hat{c}}$ と有効解像度 r_c を統合したカメラ選択尺度として、 $g_p(c) = (z_c \cdot z_{\hat{c}})^\gamma + \lambda_g r_c$ を導入

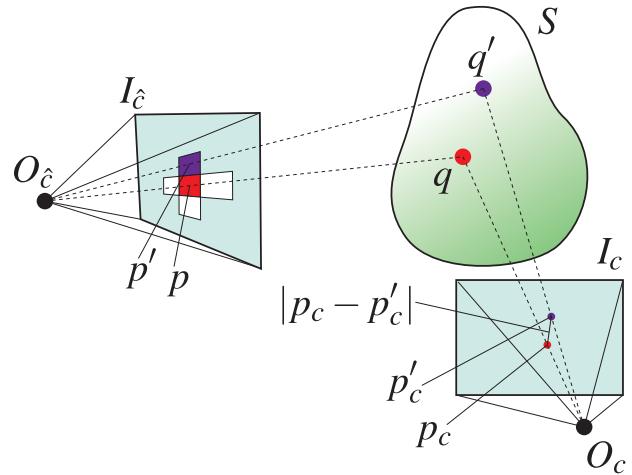


図3 仮想カメラ，対象形状，実カメラと有効解像度の関係

する。ただし、 γ は一致度の尖鋭度を示す係数 [16]、 λ_g は適当な重み係数である。

こうして定義されたカメラ選択尺度 $g_p(c)$ を用いて、各画素 p に対して $g_p(c)$ が最大となる実カメラ $\check{c} \in C_q$ を求め、この実カメラの撮影画像の画素 p_c の色を仮想カメラ画像の画素 p の色とする。以上により、多視点可変ズームカメラ環境における有効解像度を考慮した視点依存レンダリングが行われる。

3. 評価実験

3.1 入力データ

本評価実験は、以下の入力データを用いて行った。

- 撮影スタジオ：直径 4m、高さ 2.4m の十二角柱型（図4）。

- カメラ：上記のスタジオに 13 台のカメラ（Sony XCD-X710CR; XGA, 25 fps）を設置（撮影画像の例を図5, 6 に示す）。12 台の標準ズームカメラと 1 台の高ズームカメラが図4のように配置されている。

- 3次元形状（大域的最適化形状）：文献 [2] の方法を用いて生成した舞妓 1（頂点数 142,946、面数 285,912、図7(a)）と舞妓 2（頂点数 26,062、面数 52,120、図7(b)）を用いた。また、比較のために用いた Harmonized texture mapping 法の入力として、上記の形状を単純化したもの舞妓 1（頂点数 4,644、面数 9,308）、舞妓 2（頂点数 1,305、面数 2,606）を用いた。

3.2 実験および考察

本提案手法によって生成した画像を定量的に評価するため、仮想カメラ \hat{c} と実カメラ c_i のカメラパラメータを一致させ、 c_i によって撮影した画像 I_{c_i} を用いずに生成した出力画像 $I_{\hat{c}}$ による交差検定 (leave-one-out cross validation) を行って I_{c_i} との PSNR を求めた。また比較手法として、Harmonized texture mapping 法 (HTM) と、提案手法のアルゴリズム中の視点依存形状最適化を施さな

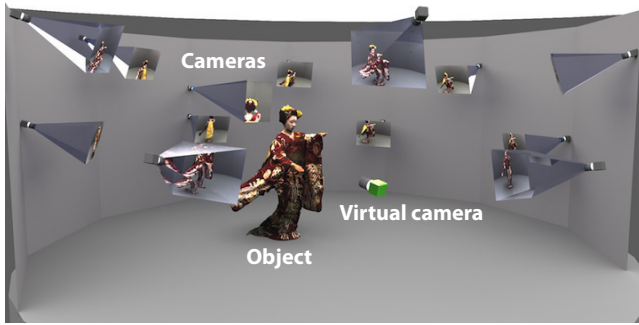


図4 撮影スタジオ

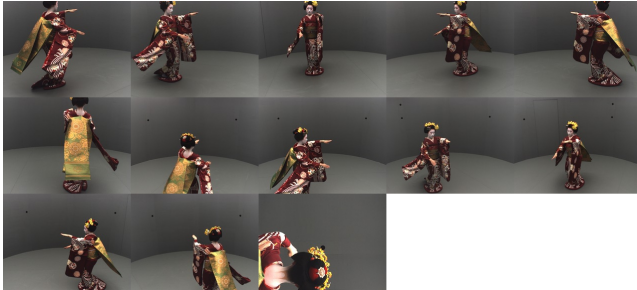


図5 入力画像(舞妓1)

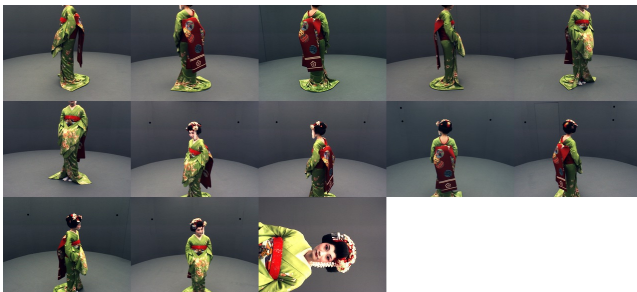
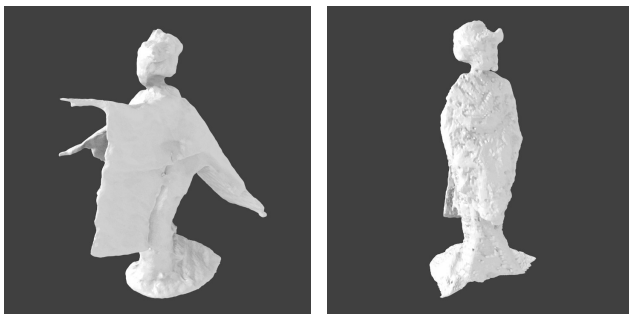


図6 入力画像(舞妓2)



(a) 舞妓1

(b) 舞妓2

図7 3次元形状(大局的最適化形状)

い方法(視点形状最適化なし)を取り上げ、それぞれについても上記の方法によって出力画像を生成した。図8, 9に求めたPSNRを、また、PSNRの計算に用いた撮影画像と出力画像の例を図10~13に示す。なお、提案手法による画像生成時のパラメータの値は以下の通りである。 $\sigma = 0.1$, $\lambda_p = 5.0$, $\lambda_d = 0.005$, $\lambda_c = 1.0$, $\lambda_g = 0.2$, $\gamma = 1.0$ 。

まず、グラフから分かることは、提案手法はHTMと比べるとPSNR野値が若干劣っていることである。そもそも、提案手法は、テクスチャ自体を変形するのではなく、形状を最適化することによって高精細な映像化を目指す方法であった。そのため、直接テクスチャの変形を行うHTMと比較すると2dB程度の差は見受けられることになるが、大きな遜色はない映像化ができていると言える。ただし注目すべきは、生成した画像の詳細を見ると、提案手法はHTMと比べてシャープな映像化ができていることである。例えば、図10の帯の模様や図12の袖の模様を見ると、HTMはローパスフィルタがかかったようになっていたが、提案手法では撮影画像とほぼ同等の解像度感で映像化できていることが分かる。これは、HTMは前処理として形状の単純化を行う必要があるが、この際に生じたわずかなずれが原因となっているのではないかと考えられる。この解像度感の向上はPSNRの値の向上として現れにくいいため、定量的に示すことが容易ではないが定性的には明らかな優位点となっている。

しかしながら、提案手法にはテクスチャの色の不連続が多く、箇所で見受けられ、これが画質の悪化を引き起こしている(例えば、図11のうなじや、図13の襟元など)。この問題については、ブレンディングによってぼかしてしまうという消極的な改善方法ではなく、高精度の形状(法線)と高精細なテクスチャが得られていることを利用した本質的な表面色(反射パラメータ)を導出するという方法で解決を目指すべきだと考えられる。この点については今後の課題としたい。

また、提案手法と形状最適化なし手法との差がそれほど現れていないが、これは、式(2)の重み係数 λ の調整や、 λ_c の問題(次節にて詳細を述べる)によって、形状の最適化が不十分となっているためだと考えられる。例えば、図13の右脇側を見ると、HTMでは帯揚げが撮影画像と同じ位置に修正できているが、提案手法では正しい位置には修正できていない。この箇所は、真の形状が凹となっているため正しく復元できていない顕著な例であるが、形状の連結性を保つ重みが強すぎて、奥行き方向に最適化できなかったためであると考えられる。

最後に、映像化にかかる時間については、CPU: Intel Core i7 (8スレッド)、メモリ: 16GBの計算機において、提案手法が2分、HTMは10分程度であった。なお、HTMについてはテクスチャ生成の前処理にかかる時間であり、一度必要なデータを生成してしまえば、仮想カメラを動かしながらの描画であっても数十ミリ秒で行えている。

4. まとめ

本論文では、視点依存形状最適化と有効解像度を考慮した視点依存テクスチャマッピングを組み合わせた、新たな自由視点画像生成法を提案した。仮想カメラ視点に依存した映像化、つまり、大局的最適化ではなく局所的

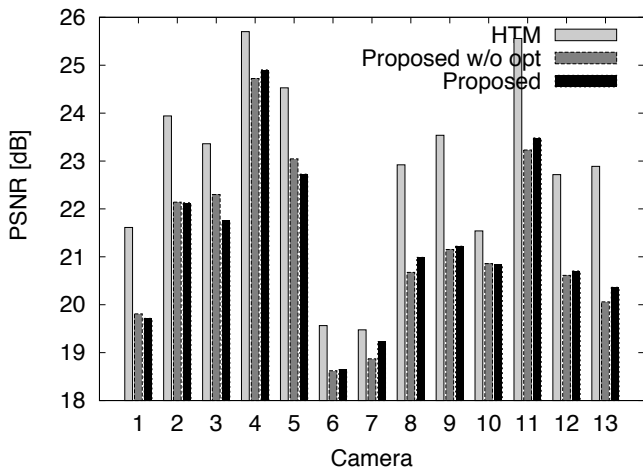


図 8 PSNR (舞妓 1)

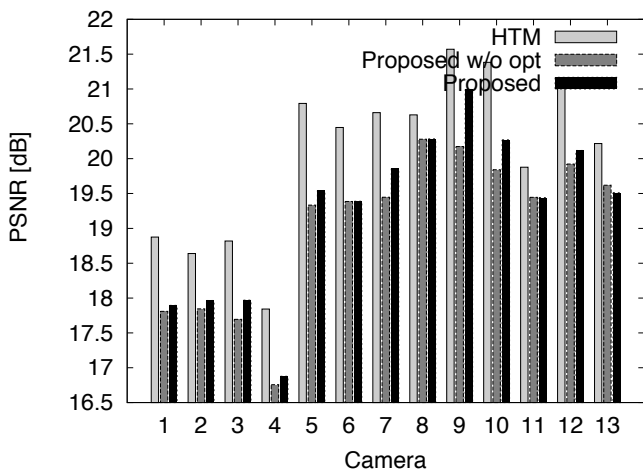


図 9 PSNR (舞妓 2)

な最適化に基づく映像化によって、より photo-consistent な形状を得ることが可能となり、さらに、高ズームのカメラによって撮影された画像を優先的に用いることによって、高精度の形状と高精細なテクスチャが同時に得られることを示した。また評価実験では、従来法との比較によって視点依存テクスチャ変形法と比べて遜色のない映像化ができることを示した。

最後に、今後取り組むべき課題として以下の2点を挙げる。

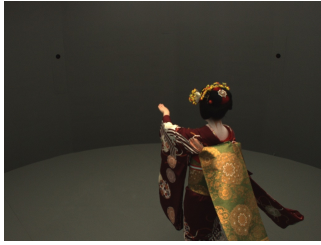
- 提案手法の現在の問題点としては、式 (2) における λ_c を一定の値としていることである。これは理論的には、画素 p の周囲の色や奥行き値を元に変化すべき値である。例えば、明らかにその画素が不連続である場合 (画像上のエッジや奥行き値が極端に変化する場合など) は λ_c を小さくしてその場所が切断されるようにし、滑らかな場合は大きくして連結を保つようにすべきである。今後は、この値の決定法についての議論が必要である。

- またもう一つの課題として、本手法により得られた高精度の形状と高精細なテクスチャを利用して、対象表面の (視点依存の) 反射パラメータ推定を行うことが挙げられる。反射パラメータの推定には、撮影時の光源

環境推定が必要となるが、3次元ビデオは人為的に環境を制御可能なスタジオ撮影を行っているので、屋外環境と比較すると単純な光源モデルによって表現できる可能性が高い。反射パラメータの推定を行えば、撮影時とは異なる照明効果の付与といった多彩な映像表現が可能となるだろう。

文 献

- [1] T. Yamaguchi, H. Yoshimoto, S. Nobuhara and T. Matsuyama: "Cell-based 3d video capture of a freely-moving object using multi-viewpoint active cameras", *IPSN Transactions on Computer Vision and Applications*, **2**, pp. 169–184 (2010).
- [2] J. Starck, A. Maki, S. Nobuhara, A. Hilton and T. Matsuyama: "The multiple-camera 3-d production studio", *IEEE Transactions on Circuit and Systems for Video Technology*, **19**, 6, pp. 856–869 (2009).
- [3] T. Kanade, P. Rander and P. J. Narayanan: "Virtualized reality: Constructing virtual worlds from real scenes", *IEEE Multimedia*, pp. 34–47 (1997).
- [4] K. N. Kutulakos and S. M. Seitz: "A theory of shape by space carving", *Proc. of ICCV*, pp. 307–314 (1999).
- [5] G. Vogiatzis, P. H. S. Torr and R. Cipolla: "Multi-view stereo via volumetric graph-cuts", *Proc. of CVPR*, pp. 391–398 (2005).
- [6] B. G. Baumgart: "A polyhedron representation for computer vision", *Proceedings of the National Computer Conference and Exposition, AFIPS '75*, pp. 589–596 (1975).
- [7] W. N. Martin and J. K. Aggarwal: "Volumetric description of objects from multiple views", *PAMI*, **5**(2), pp. 150–158 (1983).
- [8] A. Laurentini: "How far 3d shapes can be understood from 2d silhouettes", *PAMI*, **17**, 2, pp. 188–195 (1995).
- [9] S. Moezzi, L.-C. Tai and P. Gerard: "Virtual view generation for 3d digital video", *IEEE Multimedia*, pp. 18–26 (1997).
- [10] S. Nobuhara and T. Matsuyama: "Dynamic 3d shape from multi-viewpoint images using deformable mesh models", *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis*, pp. 192–197 (2003).
- [11] C. H. Esteban and F. Schmitt: "Silhouette and stereo fusion for 3d object modeling", *CVIU*, **96**, pp. 367–392 (2004).
- [12] J. Starck, A. Hilton and G. Miller: "Volumetric stereo with silhouette and feature constraints", *Proc. of BMVC*, pp. 1189–1198 (2006).
- [13] T. Tung, S. Nobuhara and T. Matsuyama: "Complete multi-view reconstruction of dynamic scenes from probabilistic fusion of narrow and wide baseline stereo", *Proc. of ICCV*, pp. 1709–1716 (2009).
- [14] D. Cremers and K. Kolev: "Multiview stereo and silhouette consistency via convex functionals over convex domains", *PAMI*, pp. 1161–1174 (2010).
- [15] M. Eisemann, B. D. Decker, M. Magnor, P. Bekaert, E. de Aguiar, N. Ahmed, C. Theobalt and A. Sellent: "Floating textures", *Eurographics '08*, **27**, 2 (2008).
- [16] T. Takai, A. Hilton and T. Matsuyama: "Harmonised texture mapping", *Fifth International Symposium on 3D Data Processing, Visualization and Transmission* (2010).
- [17] P. Felzenszwalb and D. Huttenlocher: "Efficient belief propagation for early vision", *IJCV*, **70**, pp. 41–54 (2006).



(a) 撮影画像



(b) Harmonized Texture Mapping

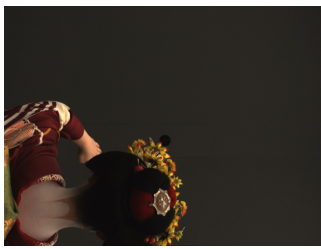


(c) 視点依存形状最適化なし



(d) 提案手法

図 10 画像生成結果 (舞妓 1, カメラ 12)



(a) 撮影画像



(b) Harmonized Texture Mapping



(c) 視点依存形状最適化なし



(d) 提案手法

図 11 画像生成結果 (舞妓 1, カメラ 13)



(a) 撮影画像



(b) Harmonized Texture Mapping

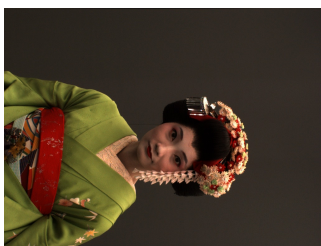


(c) 視点依存形状最適化なし



(d) 提案手法

図 12 画像生成結果 (舞妓 2, カメラ 12)



(a) 撮影画像



(b) Harmonized Texture Mapping



(c) 視点依存形状最適化なし



(d) 提案手法

図 13 画像生成結果 (舞妓 2, カメラ 13)



(a) HTM



(b) 提案手法

図 14 画像生成結果（舞妓 1，カメラ 13，拡大・回転画像）



(a) HTM



(b) 提案手法

図 15 画像生成結果（舞妓 2，カメラ 12，拡大画像）