

焦点の異なるステレオ画像からの中間視点画像の生成

濱中 智仁[†] 久保田 彰[‡] 羽鳥 好律[‡]

[†]東京工業大学大学院総合理工学研究科 〒226-8502 神奈川県横浜市緑区長津田町 4259-G2-911

E-mail: [†]t.hmnka@gmail.com, [‡]{kubota, hatori}@ip.titech.ac.jp

あらまし 近年、多視点から撮像された複数枚の画像群を用いて、任意の視点からの仮想画像(任意視点画像)を生成する研究が盛んに行われるようになった。本稿では多層の奥行からなるシーンに対して、奥行推定を経ることなく、多視点からの多焦点画像を用いて任意視点画像を生成する IBR 法を論じる。主に、多焦点画像からカメラの間の全焦点画像(中間視点画像)の生成における新しい視点内挿アルゴリズムを提案する。提案手法の大きな特徴は、従来手法[1]の入力画像の半分の枚数でほぼ同品質な生成画像を得ることができる点である。従来手法と異なり、空間不変のフィルタが直流成分で発散しない利点もある。従来手法と同様に2層の奥行からなるシーンを対象にして、実写画像を用いた実験と比較評価を行い、本手法の有効性を示す。

キーワード イメージ・ベースド・レンダリング, 中間視点画像, 空間不変フィルタ, 焦点ぼけ, ステレオ画像

View Interpolation Using Different Focus Stereo Images

Tomohito Hamanaka[†] Akira Kubota[‡] and Yoshinori Hatori[‡]

[†]Graduate School of Science and Engineering, Tokyo Institute of Technology
4259-G2-911 Nagatsuda, Midori-ku, Yokohama-shi, Kanagawa, 226-8502 Japan

E-mail: [†]t.hmnka@gmail.com, [‡]{kubota, hatori}@ip.titech.ac.jp

Abstract Recently, many approaches have been studied for generating free-viewpoint image based on real captured images. This paper presents an improved method for generating an intermediate image from differently focused stereo images for a scene of two depths. The presented method is based on linear space-invariant filtering that is independent of the scene structure. The main advantage of the presented method is to reduce the number of the required images from 4 to 2 while keeping the quality compared with our previously presented method [1].

Keyword image-based rendering, The middle aspect image, Space-invariant filter, Focus senility, Stereo image

1. はじめに

近年、複数の視点から撮像された画像群(多視点画像)を複数枚用いて、任意視点画像を生成する手法が盛んに研究されている。注目されている任意視点画像の生成方法の一つに撮像画像群の再サンプリングに基づいたイメージ・ベースド・レンダリング(IBR)法がある[2][3]。対象シーンの3次元構造が未知であっても、高品質で写実性の高い任意視点画像を生成できる特徴がある。しかし、要求される光線の取得密度は非常に高く[4]、通常のカメラでの取得は現実的に困難である場合が多い。この問題を解決するために、密な光線を補間する様々な方法が提案されてきた(たとえば[5][6])。

我々は最近の論文[1]で、焦点を変えて撮像された多視点画像から二つの段階で任意視点画像を生成する手法を示した。第一段階は撮像画像群から空間不変のフィルタ処理により、中間視点画像を密な間隔で生成する処理である。第二段階では、IBR手法の一つであるLight Field Rendering(LFR)法[7]によって、得られた密

な光線空間から容易に高品質な任意視点画像を生成する。当手法ではシーンの形状に依存しないため、一定の計算量で処理できるという特徴がある。

本稿では、第一段階の処理である密な中間視点画像の生成について新しい視点内挿の手法を提案する。近景と遠景の2層の奥行からなるシーンを対象とし、焦点はその近景と遠景にそれぞれ合わせるものとする。また直線上に等間隔に配置された複数台のカメラでシーンを撮像し、視線方向はすべてその直線に直交した奥行方向であるとする。中間視点画像を生成するために、従来手法[1]では隣接した2つの視点それぞれで近景合焦画像と遠景合焦画像の計4枚を入力画像として必要としたが、提案手法では左右一方の近景合焦画像ともう一方の遠景合焦画像の計2枚を用いて生成する点が大きく異なる。従って、従来手法では各視点で近景合焦画像と遠景合焦画像をともに撮像する必要があるのに対し、提案手法では各視点で近景合焦画像と遠景合焦画像が交互になる様に撮像すれば、必要な撮像

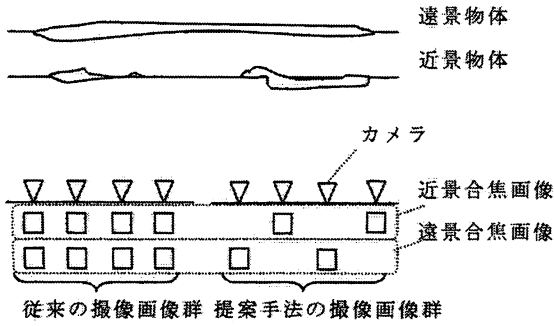


図1 従来手法と提案手法の撮像画像群の比

画像群は従来手法の半分の枚数になる。提案手法と従来手法で必要となる入力画像群の違いを図1に示す。本稿では、中間視点画像の新たな視点内挿アルゴリズムと実写画像を用いた実験結果について主に述べる。

2. 視点内挿アルゴリズム

2.1. 問題設定

左側の近景合焦画像 g_1^L と右側の遠景合焦画像 g_2^R から、中間視点画像を生成するアルゴリズムを説明する。生成画像の仮想視点は、カメラ間を $\alpha : (1-\alpha)$ に区分した点で、中間視点画像を f とする。ここで $0 \leq \alpha \leq 1$ である。目的は2枚の取得画像群 g_1^L と g_2^R から、中間視点画像 f を生成することである。この問題設定を図2に示す。

2.2. 中間視点画像のモデル

中間視点画像 f は、同じ視点から見る事ができる近景と遠景のテクスチャから構成されている。それぞれのテクスチャを f_1 と f_2 とすると、 f はそれらの和

$$f(x,y) = f_1(x,y) + f_2(x,y) \quad \dots (1)$$

として、モデル化できる。ここに、 (x,y) は画像の座標を表す。テクスチャ f_1 と f_2 の定義は

$$f_1(x,y) = \begin{cases} f(x,y) & \text{depth}(x,y) = z_1 \\ 0 & \text{depth}(x,y) = z_2 \end{cases} \quad \dots (2)$$

$$f_2(x,y) = \begin{cases} 0 & \text{depth}(x,y) = z_1 \\ f(x,y) & \text{depth}(x,y) = z_2 \end{cases} \quad \dots (3)$$

である。depth(x,y) は仮想視点における奥行き値で、近景と遠景までの距離をそれぞれ z_1 と z_2 とする。これらのテクスチャは近景と遠景を領域分割した画像であり、未知である。

2.3. 焦点の異なる2視点画像のモデル

2枚の取得画像群 g_1^L と g_2^R は(2),(3)式で定義したテクスチャを用いて近似的にモデル化できる。例えば、 g_1^L は左側の視点から近景に焦点を合わせた画像であるから、中間視点における近景と遠景のテクスチャを右方

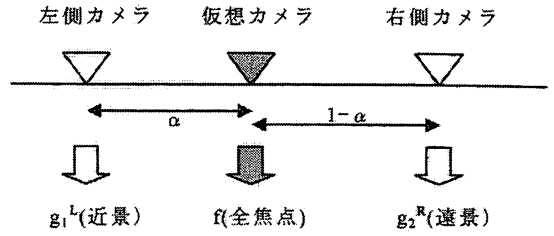


図2 問題設定。

向(x方向)に移動させた後、遠景のテクスチャにだけ焦点ぼけを与えた画像であると考えられる。視点間の近景と遠景の視差量をそれぞれ d_1 、 d_2 とすると、移動量は近景が αd_1 、遠景が αd_2 となる。焦点ぼけは、ぼけ関数 h のたたみ込みで表現できる。このぼけ関数 h はガウス関数であると仮定する。 g_2^R も同様に考えると次のようにモデル化できる。

$$\begin{cases} g_1^L(x,y) = f_1(x - \alpha d_1, y) + h_2(x,y) * f_2(x - \alpha d_2, y) \\ g_2^R(x,y) = h_1(x,y) * f_1(x + (1-\alpha)d_1, y) + f_2(x + (1-\alpha)d_2, y) \end{cases} \quad \dots (4)$$

$$h_i(x,y) = \frac{1}{\pi R_i^2} \exp\left(-\frac{x^2+y^2}{R_i^2}\right) \quad \dots (5)$$

ここに、 R_i はぼけの程度を表す量であり、ぼけの大きさはガウス関数 h の標準偏差で表される。

2.4. 周波数領域における画像生成

画像モデル(1)式、(4)式および(5)式を x と y に関しての2次元フーリエ変換すると、それぞれの次式モデルを得る。

$$\begin{cases} F(\omega_x, \omega_y) = F_1(\omega_x, \omega_y) + F_2(\omega_x, \omega_y) \quad \dots (6) \\ G_1^L(\omega_x, \omega_y) = \exp(-j\omega_x \alpha d_1) F_1(\omega_x, \omega_y) + H_2(\omega_x, \omega_y) \exp(-j\omega_x \alpha d_2) F_2(\omega_x, \omega_y) \\ G_2^R(\omega_x, \omega_y) = H_1(\omega_x, \omega_y) \exp(j\omega_x (1-\alpha)d_1) F_1(\omega_x, \omega_y) + \exp(j\omega_x (1-\alpha)d_2) F_2(\omega_x, \omega_y) \end{cases} \quad \dots (7)$$

$$H_i(\omega_x, \omega_y) = \exp\left\{-\frac{R_i^2}{4}(\omega_x^2 + \omega_y^2)\right\} \quad \dots (8)$$

ここで大文字の画像は、小文字の画像の2次元フーリエ変換を表し、 (ω_x, ω_y) は (x,y) に対する角周波数、 j は虚数である。(7)式は F_1 と F_2 の連立方程式となっている。(7)式より F_1 と F_2 についてそれぞれ解いて、(6)式代入し、整理すると次式を得る。

$$F(\omega_x, \omega_y) = K^L(\omega_x, \omega_y) G_1^L(\omega_x, \omega_y) + K^R(\omega_x, \omega_y) G_2^R(\omega_x, \omega_y) \quad \dots (9)$$

K^L と K^R は、各取得画像に施すべきフィルタの周波数特性であり、次式となる。ここで K^L 、 K^R 、 H_1 、 H_2 の (ω_x, ω_y) は略する。

$$\begin{cases} K^L = \frac{\begin{bmatrix} \exp(j\omega_x \alpha d_1) \\ -H_1 \exp(j\omega_x (d_1 - (1-\alpha)d_2)) \\ 1 - H_1 H_2 \exp(j\omega_x (d_1 - d_2)) \end{bmatrix}}{\begin{bmatrix} \exp(-j\omega_x (1-\alpha)d_2) \\ -H_2 \exp(j\omega_x (\alpha d_1 - d_2)) \\ 1 - H_1 H_2 \exp(j\omega_x (d_1 - d_2)) \end{bmatrix}} \\ K^R = \frac{\begin{bmatrix} \exp(j\omega_x \alpha d_1) \\ -H_1 \exp(j\omega_x (d_1 - (1-\alpha)d_2)) \\ 1 - H_1 H_2 \exp(j\omega_x (d_1 - d_2)) \end{bmatrix}}{\begin{bmatrix} \exp(-j\omega_x (1-\alpha)d_2) \\ -H_2 \exp(j\omega_x (\alpha d_1 - d_2)) \\ 1 - H_1 H_2 \exp(j\omega_x (d_1 - d_2)) \end{bmatrix}} \end{cases} \quad \dots(10)$$

さらに分母の実数化し、次式を得る。

$$\begin{cases} K^L = \frac{\begin{bmatrix} \exp(j\omega_x \alpha d_1) \\ -H_1 \exp(j\omega_x (d_1 - (1-\alpha)d_2)) \\ -H_1 H_2 \exp(j\omega_x (-(1-\alpha)d_1 + d_2)) \\ +H_1^2 H_2 \exp(j\omega_x \alpha d_2) \end{bmatrix}}{1 + H_1^2 H_2^2 - 2H_1 H_2 \cos\{\omega_x (d_1 - d_2)\}} \\ K^R = \frac{\begin{bmatrix} \exp(-j\omega_x (1-\alpha)d_2) \\ -H_2 \exp(j\omega_x (\alpha d_1 - d_2)) \\ -H_1 H_2 \exp(j\omega_x (d_1 - \alpha d_1)) \\ +H_1 H_2^2 \exp(-j\omega_x (1-\alpha)d_1) \end{bmatrix}}{1 + H_1^2 H_2^2 - 2H_1 H_2 \cos\{\omega_x (d_1 - d_2)\}} \end{cases} \quad \dots(11)$$

(9)式は周波数領域において、2枚の取得画像群 G_1^L と G_2^R に対して、 K^L と K^R によるフィルタリングを行った結果を加算することによって、中間視点画像 f を生成できることを示している。図3に(9)式に基づいた提案手法の流れを示す。二つの入力画像をフーリエ変換し、それぞれフィルタとの積をとる。得られた結果を加算して、逆フーリエ変換をすることによって、中間視点画像 f が生成される。従来手法と同様に、用いるフィルタは(10)式に基づいてぼけ関数と視差量だけで決定される。従って、提案手法も奥行画像の推定や領域分割を行う必要はなく、どの領域が近景であるかといった情報にまったく依存しない。

しかしながら、(10)式の分母に注目すると、 $(\omega_x, \omega_y) = (0, 0)$ の直流においてのみ、フィルタの分母値が零となることがわかる。各フィルタの直流成分付近の極限をとると、

$$\begin{cases} \lim_{\omega_y \rightarrow 0} \lim_{\omega_x \rightarrow 0} K^L(\omega_x, \omega_y) = \frac{R_1^2}{(R_1^2 + R_2^2)} \\ \lim_{\omega_y \rightarrow 0} \lim_{\omega_x \rightarrow 0} K^R(\omega_x, \omega_y) = \frac{R_2^2}{(R_1^2 + R_2^2)} \end{cases} \quad \dots(12)$$

$$\begin{cases} \lim_{\omega_x \rightarrow 0} \lim_{\omega_y \rightarrow 0} K^L(\omega_x, \omega_y) = 1 - \alpha \\ \lim_{\omega_x \rightarrow 0} \lim_{\omega_y \rightarrow 0} K^R(\omega_x, \omega_y) = \alpha \end{cases} \quad \dots(13)$$

となるので、実数値に収束し、発散はしない。従来手法ではフィルタの直流成分が発散していたのに対し、提案手法では発散しないため、直流成分付近を特に補正する必要はない。ただし(12)、(13)式より R_1 や R_2 、 α を変えると極限值が異なる。低水平周波数で視差量を減少させると(13)式の左辺が(12)式の値と一致するため、実際の直流での値は(12)式の値を用いる。実際に視点間距離が 8[mm]の場合、3次元上にプロットしたフィルタ K^L を図4に示す。 α の値を 0.2、0.5、0.8と変えて、フィルタの実部と虚部をそれぞれ上から交

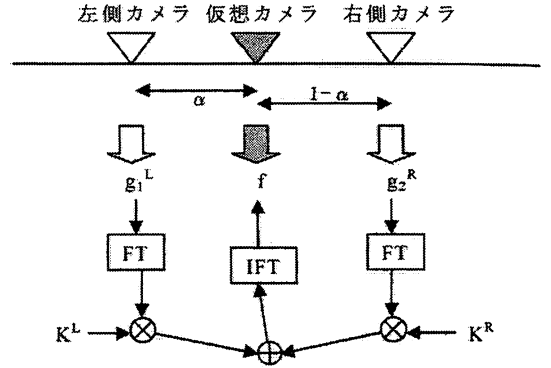


図3 提案手法：フィルタによる中間視点画像の生成方法

互に示している。左側がフィルタの全体図で、右側が直流成分付近の拡大図である。いずれのフィルタも水平周波数成分から見て、正弦波の形をしている。また発散はしていないことが全体図より確認できる。 α が0.5(中点)より離れるほど(12)、(13)式の直流成分値が異なるため、直流成分付近で拡大図の様にねじれることもわかる。全体図より α が増えれば、振動数(波数)が増え、周期が短くなる特徴がある。逆に α が減れば、振動数も減り、周期が長くなる。またフィルタ K^R の場合は α が増えるほど、振動数は減り、周期は長くなる。つまり、カメラの仮想視点位置が左側カメラに近づけば、フィルタ K^L の振動数が減り、フィルタ K^R の振動数が増える。逆にカメラの仮想視点位置が右側カメラに近づけば、フィルタ K^L の振動数が増え、フィルタ K^R の振動数が減る。

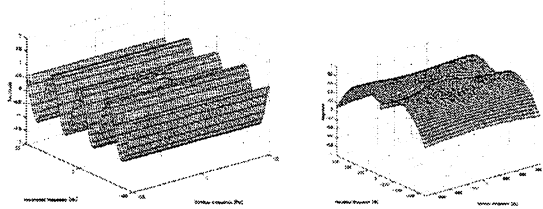
3. 実験と考察

3.1. 中間視点画像の生成環境

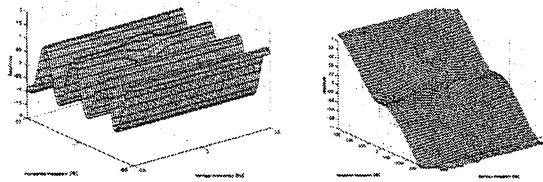
水平な直線上を 2[mm]間隔で9視点、交互に近景と遠景に焦点を合わせて撮像した画像群を用いて、中間視点画像の生成を行った。シーンは近景物体(ペンが入ったコップ)と本の背景からなり、それぞれ奥行はおおよそ 1000[mm]と 1600[mm]にある。画像サイズは 280×256画素である。既存の推定方法[1]により、ぼけの大きさ(ガウス関数の標準偏差 R_1, R_2)を順に $R_1=R_2 = 3.6$ [pixel]とした。また左右のカメラ視点間が 8[mm]の場合、近景合焦画像と遠景合焦画像の視差量を順に、 $d_1=18.5$ [pixel], $d_2 = 11.4$ [pixel]とした。

3.2. 視点間距離の異なる画像の比較

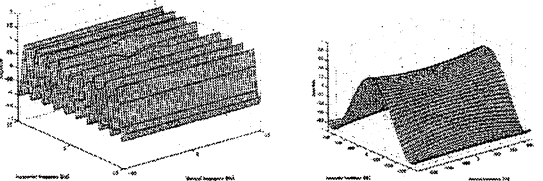
カメラの視点間距離が 8[mm]と 16[mm]の場合で、中間視点画像($\alpha=0.5$)をそれぞれ生成した。4枚の入力画像を用いる従来手法と合わせて、実験結果を図5に示す。また、視点間距離が 8[mm]の場合の焦点の異なる



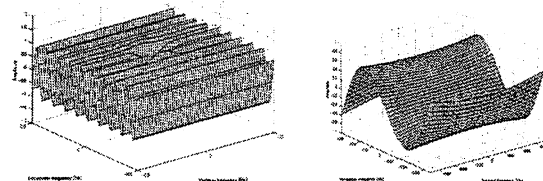
(a) $\text{Re}\{K^L\}$ $\alpha:0.2$



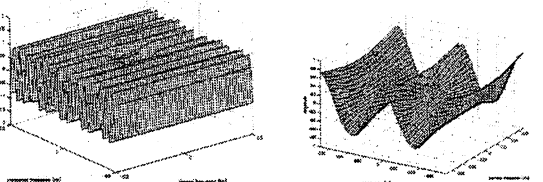
(b) $\text{Im}\{K^L\}$ $\alpha:0.2$



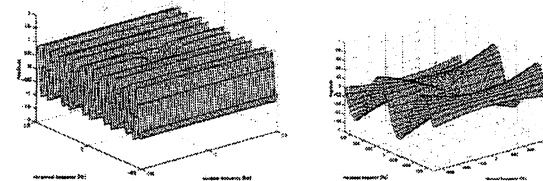
(c) $\text{Re}\{K^L\}$ $\alpha:0.5$



(d) $\text{Im}\{K^L\}$ $\alpha:0.5$



(e) $\text{Re}\{K^L\}$ $\alpha:0.8$

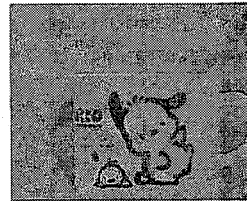


(f) $\text{Im}\{K^L\}$ $\alpha:0.8$

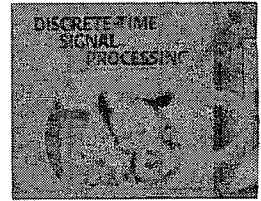
図4 フィルタの周波数特性

入力画像2枚と、実際に同じ中間視点で絞りを小さくして撮像した正解画像も示している。提案手法と従来手法を見比べると、視点間距離が8[mm]の場合は比較的劣化なく良好な中間視点画像が生成できることが確認できる。一方、視点間距離を16[mm]まで広げると、提案手法では残像がやや目立つ。この残像による中間視点画像の品質への影響については3.4節で定量的に考察する。

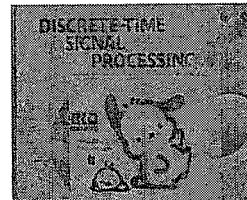
図5の(c)と(d)の一部分(ペンとコップと背景の文字が重なっている部分)を拡大した図を図6に示す。見比べると、背景の「G」の文字が提案手法では従来手法より劣化していることがわかる。提案手法ではオクルージョンの部分となるため本質的に解決はできない。左側の遠景合焦画像を用いる従来手法でも、領域分割



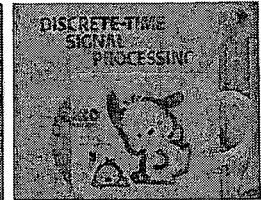
(a) 入力画像: g_1^L



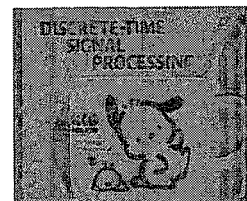
(b) 入力画像: g_2^R



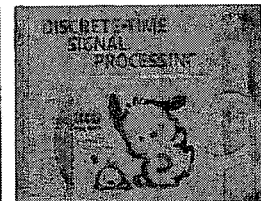
(c) 提案手法:8[mm]



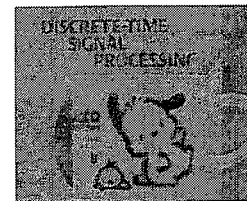
(d) 従来手法:8[mm]



(e) 提案手法:16[mm]

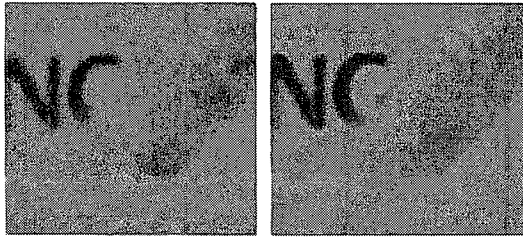


(f) 従来手法:16[mm]



(g) 正解画像:f

図5 中間視点画像の生成結果



(a) 提案手法 (b) 従来手法
図 6 オクルージョン部分の劣化

を行わないため、境界近傍を高精度に画像生成することは困難な場合が多い。

3.3. 仮想視点位置の異なる画像の比較

図 7 に図 5 の 2 枚の入力画像から提案手法で生成した中間視点画像の例を示す。仮想カメラの異なる視点位置(α の値で 0, 0.2, 0.4, 0.6, 0.8, 1 に相当する位置)で生成された中間視点画像とその一部の拡大画像を示している。近景と遠景の領域とも焦点の合った画像が生成されている。また、近景と遠景の領域の相対的な位置関係から、視点位置に応じた視差が生成されていることも確認できる。近景と遠景の境界領域においては、画像モデルが正しくないのにも関わらず、オクルージョン部分を除いて目立った劣化は生じていない。以上の観測から、提案手法も従来手法と同様に良好な中間視点画像が生成できていることがわかる。精度については、次節で考察する。

3.4. 精度の評価と考察

中間視点画像の生成手法の精度を評価するために次のような追加実験を行った。3.2.節と同実験環境で、カメラの視点間距離が 4[mm]、12[mm]の場合の中間視点画像を生成した。よって、提案手法と従来手法それぞれ、視点間距離が 4[mm]、8[mm]、12[mm]、16[mm]の場合の計 8 枚の生成画像が得られた。得られた 8 枚の生成画像はいずれも図 5 の正解画像(g)と同じ視点位置である。そこで、この 8 枚の生成画像と正解画像(g)の平均自乗誤差(MSE)を用いて、生成手法の精度を定量的に評価する。

提案手法と従来手法を用いて生成した画像の MSE をプロットした結果を図 8 に示す。視点間距離(横軸)に対する MSE(縦軸)を示している。一般的に視点間距離が広がれば MSE も増加する。またどちらも 12[mm]までは比較的安定である。いずれの場合も提案手法は従来手法に比べて、ほぼ精度は同等であることが確認できる。しかし、図 5 の(c)、(f)より視点間距離が 16[mm]になると残像が観測される。その要因の一つには、周

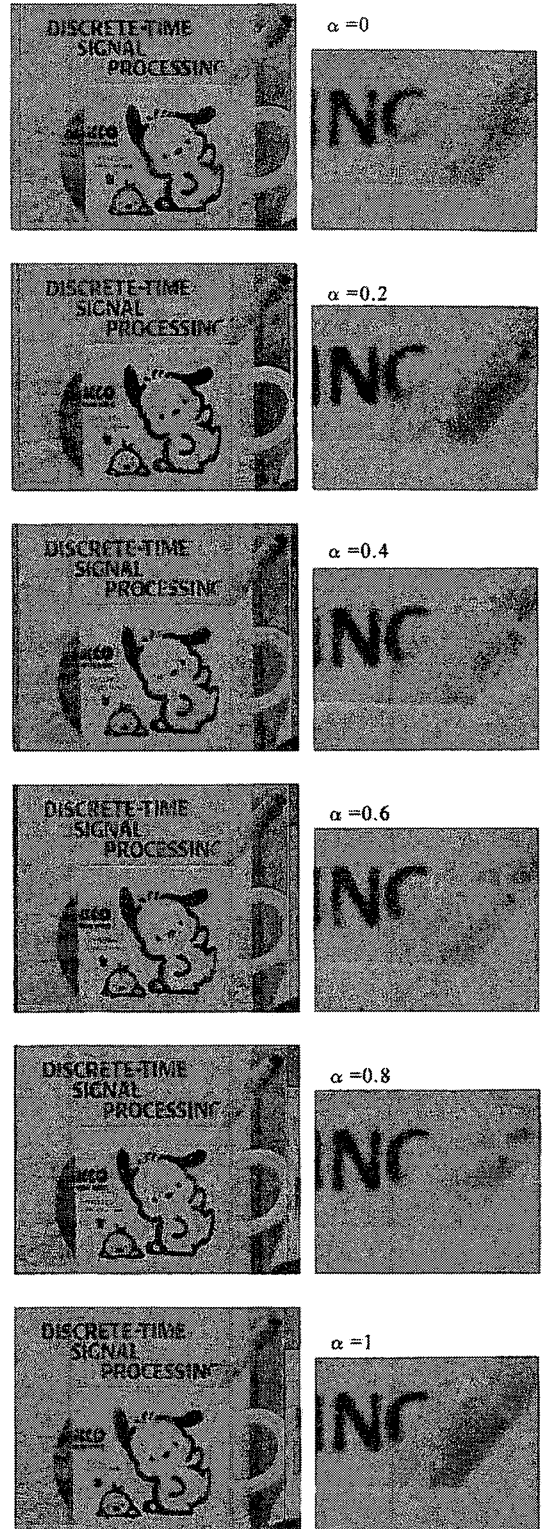


図 7 中間視点画像の生成結果

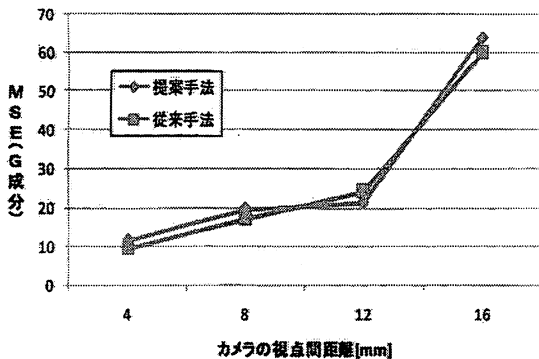


図8 提案手法と従来手法を用いて生成した中間視点画像の平均自乗誤差(MSE)

波数フィルタの理論式((11)式)の分母に \exp を含んでいるため、従来手法に比べて観測誤差を強調する可能性がある。

最後に仮想視点位置の異なる画像の評価を行った。カメラの視点間距離を 8[mm]に固定し、 α を 0, 0.25, 0.5, 0.75, 1 と変えて、提案手法と従来手法それぞれを用いて中間視点画像を生成した。それぞれの仮想カメラ視点ごとに撮影した正解画像を用いて、同様に各 MSE を求めた。結果は、いずれの場合も提案手法と従来手法の精度差は 10 以下であり、ほぼ同等であった。

4. おわりに

本論文では、2 層の奥行からなる対象シーンに対して、中間視点画像を従来手法の半分の撮像画像から生成する新しい視点内挿アルゴリズムを提案した。多視点の各点において、交互に近景合焦画像と遠景合焦画像を撮像し、取得した画像群から各視点間の中間視点画像を生成する手法を提案した。中間視点画像生成の従来手法は、左右それぞれの近景合焦画像と遠景合焦画像の計 4 枚を必要としたのに対し、提案手法では一方を近景に、もう一方を遠景に焦点を合わせた左右 2 枚の撮像画像から生成できる。また提案手法は、次の点において従来手法と同等な性質を持つ。周波数フィルタは空間不変なため、領域分割や奥行画像の推定を行う必要はなく、中間視点画像を生成できる。フィルタは、近景と遠景のぼけ量と視差量で決定されるため、画像生成に必要な計算量は一定で、物体の形状に依存しない。中間視点画像の品質は視点間距離が広いと観測誤差による残像が生じるが、定量的にはほとんど劣化なく、安定的に画像生成ができることを示した。

今後の課題として、提案手法で生じる残像の原因の分析が必要である。原因の一つとして考えられるぼけ関数の理想化による誤差を確認する方法として、入力

画像や正解画像を CG で作成し、観測誤差のない環境下における生成手法の精度評価比較が挙げられる。また、CG 画像を用いて、人工的な観測誤差が中間視点画像に与える影響実験が挙げられる。

文 献

- [1] 久保田, 相澤, “多視点からの多視点画像を用いた密な光線空間の構築とそれに基づいた任意視点画像の生成,” 映情学誌, **60**, 5, pp. 797-806, (2006)
- [2] H-Y. Shum, S. B. He, and S-C. Chan, “Survey of Image-Based Representations and Compression Techniques”, *IEEE Trans. On Circuit and System for Video Technology*, **13**, 11, pp. 1020 – 1037, (2003)
- [3] C. Zhang and T. Chen, “A survey on image-based rendering – representation, sampling and compression,” *EURASIP Signal Processing: Image Communication*, **19**, pp. 1-28, (2004)
- [4] J-X. Chai, X. Tong, S.-C. Chan, H.-Y. Shum, “Pleoptic Sampling,” in proc. *ACM SIGGRAPH'00*, pp. 307-318, (2000)
- [5] 高橋, 久保田, 苗村, “Light Field Rendering における合焦判定と全焦点画像合成,” 信学論, **J88-D-II**, 3, pp. 573-584 (Mar. 2005)
- [6] 中西, 藤井, 木本, 谷本, “EPI 上の対応点軌跡を用いた適応フィルタによる光線空間データ補間,” 映情学誌, **56**, 8, pp. 1321-1327, (2002)
- [7] M. Levoy and P. Hanrahan, “Light field rendering”, in proc. of *ACM SIGGRAPH'96*, pp. 31-42 (1996)
Survey on stereo matching