

修士論文要旨 (2021 年度)

ステレオ画像に対するシーン非依存の最適視点内挿 フィルタ

Scene-Independent Optimum View Interpolation Filters for Stereo Images

電気電子情報通信工専攻 中坪 征二

1 はじめに

視点内挿とは、多視点画像を用いて中間視点画像を生成する問題である。この問題の解法として、多視点画像間の対応点を推定し、対応点の画素値を混合する方法が主に検討されている [1], [2]。

本稿では、ステレオ画像に対してフィルタによる視点内挿法を提案する。提案手法では、各画像をフィルタに通した後、加算することによって中間視点画像を生成する。フィルタは、最小二乗の意味で中間視点画像を最良近似するように設計される。フィルタは空間不変であるため、シーンに依存しない。すなわち、提案手法は対応点推定が不要である。

筆者は、2層からなるシーンに対して、4視点画像から中間視点画像を内挿する空間不変フィルタを提案した [3]。しかし、フィルタは反復近似であり、対象シーンの奥行は2層に限られていた。3層以上からなる一般のシーンに対しては、反復再構成 [4] が論じられてきたが、中間視点画像を直接生成するフィルタの導出は検討されていない。

また、高橋は、ステレオ画像に対して、中間視点画像の二乗誤差を最小にする最適フィルタを導出している [5]。このフィルタは、事前に推定した対応点に誤りがあったとしても、その影響を最小にすることができる。提案手法では、対応点の推定は行わないため、対応点誤りが比較的大きい場合についての最適フィルタを論じていることになる。

2 視点内挿フィルタ

2.1 問題の記述

ステレオ画像を左視点画像 $g_L(x, y)$ と右視点画像 $g_R(x, y)$ とする。中間視点の位置は、基線を $\alpha : (1-\alpha)$ に内分した点とし、中間視点画像を $f(x, y)$ とする。本稿

の目的は $g_L(x, y)$ と $g_R(x, y)$ から、中間視点画像 $f(x, y)$ を

$$f(x, y) = k_L(x, y) * g_L(x, y) + k_R(x, y) * g_R(x, y) \quad (1)$$

により内挿するフィルタ $k_L(x, y)$ と $k_R(x, y)$ を導出することである。ここに、 $*$ は二次元畳み込み積分を表す。導出は周波数領域で行い、フィルタの周波数特性 $K_L(\xi, \eta)$ と $K_R(\xi, \eta)$ を導出する。 (x, y) は画像の空間座標、 (ξ, η) は空間角周波数を表すが、以後、必要な場合のみ明示することにする。

2.2 画像の取得モデル

シーンが N 層のレイヤから構成されているとする。レイヤの奥行は、その視差が一定間隔で変化するように設定しておく。各レイヤが中間視点で取得する画像をレイヤ画像と呼び、 $\phi_n (n = 1, 2, \dots, N)$ と表す。レイヤ画像を用いて、中間視点画像 f をそれらの和としてモデル化する。

$$f = \sum_{n=1}^N \phi_n \quad (2)$$

右視点画像と左視点画像は、各レイヤ画像を水平に移動させた画像の和としてモデル化する。

$$\begin{cases} g_L = \sum_{n=1}^N \phi_n (x - \alpha d_n) \\ g_R = \sum_{n=1}^N \phi_n (x + (1 - \alpha) d_n) \end{cases} \quad (3)$$

ここに、 d_n は ϕ_n のステレオ画像間の水平視差を表す。式 (2) と式 (3) を二次元フーリエ変換し、周波数領域におけるモデルをそれぞれ次のように得る。

$$F = 1_N^T \phi \quad (4)$$

$$g = H\phi \quad (5)$$

ここに、 T は転置を表す。は、すべての要素が1の N 次元ベクトル、 g 、 ϕ および H は以下のとおりである。

$$\begin{aligned} g &= (G_L, G_R)^\top \\ \phi &= (\Phi_1, \Phi_2, \dots, \Phi_N)^\top \end{aligned} \quad (6)$$

$$H = \begin{pmatrix} W^{-\alpha d_1} & W^{-\alpha d_2} & \dots & W^{-\alpha d_N} \\ W^{(1-\alpha)d_1} & W^{(1-\alpha)d_2} & \dots & W^{(1-\alpha)d_N} \end{pmatrix}$$

ただし、 $W \triangleq \exp(j\xi)$

大文字の関数は対応する小文字の関数の二次元フーリエ変換を表す。

2.3 視点内挿フィルタ

式 (5) から、 H の疑似逆行列 H^+ を用いて、レイヤ画像 ϕ の最小二乗解 H^+g を求める。これを式 (4) の右辺に代入することにより、ステレオ画像 g から中間視点画像 F を推定する式が得られる。

$$\hat{F} = 1_N^T H^+ g$$

したがって、内挿フィルタの周波数特性 K_L と K_R は、 $1_N^T H^+$ の要素として

$$(K_L, K_R) = 1_N^T H^+ \quad (7)$$

と表される。

H の特異値分解を用いて疑似逆行列 H^+ を表現し、その結果を式 (7) に代入し、整理すると

$$(K_L, K_R) = \sum_{m=1}^2 \frac{\mu_m}{\lambda_m} u_m^* \quad (8)$$

を得る。ここに、 $*$ は共役転置を表す。 λ_m は、

$$HH^* = \begin{pmatrix} N & ND_N(\Delta\xi)W^{-d_c\xi} \\ ND_N(\Delta\xi)W^{d_c\xi} & N \end{pmatrix}$$

の固有値であり、

$$\lambda_m = N \{(-1)^{m-1} |D_N(\Delta\xi)| + 1\} \quad (9)$$

となる。固有値は非負の実数となり、 $\lambda_1 > \lambda_2 > 0$ の関係が成り立つ。ただし、 $\Delta \triangleq (d_1 - d_N)/(N - 1)$ および $\Delta_c \triangleq (d_1 + d_N)/2$ であり、それぞれの視差の間隔と中心を表す。 $D_N(\Delta\xi)$ は次式の *Dirichlet* 関数 (*aliased sinc* 関数) から計算される。

$$D_N(x) \triangleq \begin{cases} \frac{\sin(xN/2)}{N \sin(x/2)} & x \neq 2\pi l \quad (l = 0, \pm 1, \pm 2, \dots) \\ (-1)^l(N-1) & x = 2\pi l \quad (l = 0, \pm 1, \pm 2, \dots) \end{cases} \quad (10)$$

u_m は固有値 λ_m に属する HH^* の固有ベクトルであり、

$$u_m = \frac{1}{\sqrt{2}} \left((-1)^{m-1} \text{sgn} \{D_N(\Delta\xi)\} W^{-d_c}, 1 \right)^\top \quad (11)$$

となる。 λ_m と u_m は、中間視点の位置 α に依存しないことがわかる。これは、 HH^* が α に依存しないことによる。

μ_m は、次式のように計算され、中間視点の位置 α に依存した値となる。

$$\begin{aligned} \mu_m &= \langle \mathbf{1}_N, H^* u_m \rangle = \langle H \mathbf{1}_N, u_m \rangle \\ &= \frac{N}{\sqrt{2}} \left\{ (-1)^{m-1} D_N(\alpha\Delta\xi) \text{sgn} \{D_N(\Delta\xi)\} \right. \\ &\quad \left. + D_N((1-\alpha)\Delta\xi) \right\} W^{-(1-\alpha)d_c} \end{aligned}$$

以上の結果を式 (8) に代入すると、フィルタ K_L と K_R は次式のようになる。

$$\begin{cases} K_L = \frac{D_N(\alpha\Delta\xi) - D_N(\Delta\xi)D_N((1-\alpha)\Delta\xi)}{1 - D_N^2(\Delta\xi)} W^{\alpha d_c} \\ K_R = \frac{D_N((1-\alpha)\Delta\xi) - D_N(\Delta\xi)D_N(\alpha\Delta\xi)}{1 - D_N^2(\Delta\xi)} W^{-(1-\alpha)d_c} \end{cases}$$

2.4 奥行が連続の場合の視点内挿フィルタ

基線長と対象シーンの奥行範囲の比が固定であれば、視差の範囲である $d_1 - d_N$ は一定となる。 $d_r \triangleq d_1 - d_N$ とおく。 d_r を一定として、層数 N を無限にすると、 $D_N(\delta d\xi)$ は次の *sinc* 関数に収束する。

$$\lim_{N \rightarrow \infty} D_N(\Delta d\xi) = \lim_{N \rightarrow \infty} D_N \left(\frac{d_r}{N-1} \xi \right) = \text{sinc} \left(\frac{d_r \xi}{2\pi} \right)$$

これにより、 N を無限としたとき、フィルタ K_L と K_R は次式のように得られる。

$$\begin{cases} K_L = \frac{\text{sinc}(\frac{\alpha d_r \xi}{2\pi}) - \text{sinc}(\frac{d_r \xi}{2\pi}) \text{sinc}(\frac{(1-\alpha)d_r \xi}{2\pi})}{1 - \text{sinc}^2(\frac{d_r \xi}{2\pi})} \cdot \exp \{j\xi \alpha d_c\} \\ K_R = \frac{\text{sinc}(\frac{(1-\alpha)d_r \xi}{2\pi}) - \text{sinc}(\frac{d_r \xi}{2\pi}) \text{sinc}(\frac{\alpha d_r \xi}{2\pi})}{1 - \text{sinc}^2(\frac{d_r \xi}{2\pi})} \cdot \exp \{-j\xi (1-\alpha) d_c\} \end{cases}$$

フィルタの分母は直流である $\xi = 0$ において零点をもつが、フィルタの直流への極限值は以下のように存在する。

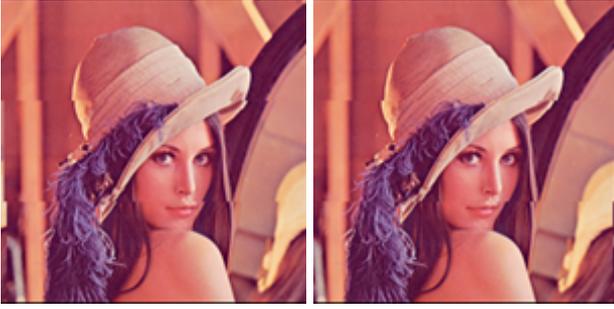
$$\lim_{\xi \rightarrow 0} K_L = 1 - \alpha, \quad \lim_{\xi \rightarrow 0} K_R = \alpha$$

したがって、フィルタは全ての周波数で存在することがわかる。

3 シミュレーション

3.1 ステレオ画像の準備

ステレオ画像はモデル式 (3) を用いて合成した。レイヤ画像 $\phi_1, \phi_2, \dots, \phi_N$ は、画像 “Lenna” (256 × 256



(a) 左視点画像 g_L (b) 右視点画像 g_R

図 1: モデルとして作成したステレオ画像

画素)を垂直方向に N 等分して、画像の上から順に切り出して作成した。ただし、画像の推移は巡回推移とした。本実験では、中点 ($\alpha = 0.5$) における画像が画像 “Lenna” と同一となるように設定した。また、 N 番目のレイヤが無遠方にあるとして $d_N = 0$ とした。レイヤ数を $N = 3$ 、視差の範囲を $d_1 = 10$ としたときのステレオ画像の例を図 1 に示す。画像の上から $1/3$ の領域 (レイヤ画像 ϕ_1) は、水平視差が 10 画素になっている。画像の下から $1/3$ の領域 (レイヤ画像 ϕ_4) は、視差はなく、同一の画像になっている。

3.2 再構成画像の結果と評価

レイヤ数が 3 枚であるときのステレオ画像を用いて、視点が中心 ($\alpha = 0.5$) である中間視点画像を再構成し、品質の評価を行った。品質の評価には PSNR (Peak Signal to Noise Ratio) を用いた。評価に用いる正解画像は、中間視点画像のモデル式 (2) を用いて作成した。また、左右の画像の平均をとって中間視点画像を抽出する averaging method を比較対象にして実験を行った。視差を $d_1 = 2, 4, 6, 8$ および 10 と変化させた場合の結果を表 1 に示す。視差 d_1 を大きくすると、方法に関係なく品質が低下することがわかる。これは、視差が大きくなるとフィルターの周波数特性の零点が増加するため、それらの周波数成分の復元ができなくなるためである。また、結果は提案手法が視差の範囲に関係なく、averaging method よりも制度が高いことがわかる。図 2 は、視差 d_1 を変化させた際に、 α を 0.1 から 0.9 に 0.1 間隔で変更した場合の再構成画像の PSNR を示す。視点が中心に近いほど PSNR が低くなっていることがわかる。視点がステレオ画像の視点に近い場合、近い画像のフィルターのゲインが大きくなるため品質が高くなると考えられる。

2 つの手法の再構成画像と正解画像の絶対差分画像を図 3 と 4 に示す。視差 d_1 を 2 と 10 に変更すると、視差が大きいほど誤差が大きくなることがわかる。ただ

表 1: $PSNR(N = 3)[dB]$

視差	$d_1 : 5$	$d_1 : 4$	$d_1 : 6$	$d_1 : 8$	$d_1 : 10$
空間不変フィルタ	40.5	31.5	28.1	26.5	25.6
averaging method	39.5	30.2	26.6	24.9	23.8

し、提案手法のエッジは、averaging method のエッジよりも薄くなっていることがわかる。

4 まとめ

本論文では、ステレオ画像から中間視点画像を補間するための最適な空間不変フィルタを導出した、シミュレーション結果は、提案手法が averaging method よりも高品質な中間視点画像を再構成することを示した。今後は実像を使った実験を行う。

参考文献

- [1] S.C. Chan, H.Y. Shum, and K.T. Ng, “Image-based rendering and synthesis,” *IEEE Signal Process. Mag.*, vol.24, no.7, pp.22-33, 2007.
- [2] Aljoscha Smolic, “3D video and free viewpoint video- From capture to display,” *Pattern Recognition*, vol.44, no.9, pp.1958-1968, 2011.
- [3] A. Kubota, K. Aizawa and Tsuhan Chen, “Direct filtering method for image based rendering,” *IEEE International Conference on Image Processing 2005*, 2005, pp. III-1, doi: 10.1109/ICIP.2005.1530313.
- [4] A. Kubota, K. Takahashi, K. Aizawa, T. Chen, “All Focused Light Field Rendering,” *proc. of EGSR2004*, pp.235-242, 2004.
- [5] K. Takahashi, “Theoretical analysis of view interpolation with inaccurate depth information,” *IEEE Trans. on Image Process.*, vol.21, no.2, pp.718-732, 2012.

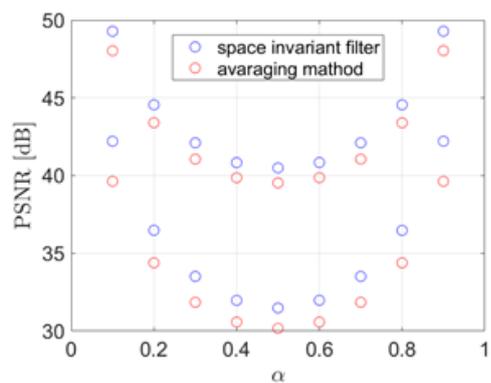
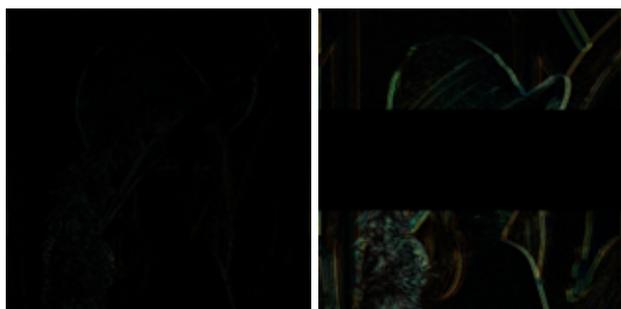
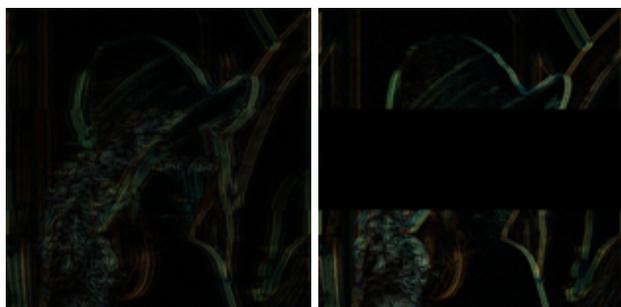


図 2: 様々な視点からの再構成された中間視点画像の PSNR



(a) 空間不変フィルタ (b) averaging method

図 3: 絶対差分画像 ($d_1 = 2$)



(a) 空間不変フィルタ (b) averaging method

図 4: 絶対差分画像 ($d_1 = 10$)