

3Dモデル交換方式を用いた携帯電話端末による物体認識技術の研究（継続）

代表研究者 羽 鳥 好 律 東京工業大学大学院総合理工学研究科教授
 共同研究者 久 保 田 彰 東京工業大学大学院総合理工学研究科助教

1 はじめに

携帯電話端末に、各個人ごとにカスタマイズされたネットワークヒューマンインタフェースを提供する「情報介助装置」としての役割を担わせることができれば、結果として、過大なりテラシの習得を求めることなく、高齢者・幼児・日本滞在者等も含め多くの人々が、ユビキタス社会の恩恵に浴することが出来るようになることが期待される。

本研究では、携帯電話端末と言う比較的限られた演算能力を有する端末に、他の機器との相互情報交換の機能とある程度のインテリジェントな情報処理機能を負わせることで、例えば、不慣れな機器でも利用者個人を同定した上で普段使い慣れている自己の携帯端末を介して操作を実行できるような仕組みを考える。最終的にはその為に必要となる、携帯端末でのモデル交換ベースでの簡易な物体認識機能の実現とデータ交換方式が有用となる要件を明らかにする。特に今年度においては、それを実現可能とするためのコア要素技術として、端末により取得した複数画像群より自由視点画像を再構成し、与えられた交換モデルの不足分を補う Light field rendering (LFR) [1]手法について検討する。

LFR は複数の視点で得られた画像群から自由視点画像を再構成する基本的な方法である。元の構造情報を陽に用いることのない方法なので、不十分なモデル情報からも、計算処理で自由視点画像を再構成することが可能である。但し、最初の段階において、対象物体の存在する位置を 1 focal plane 上に存在すると仮定するところから計算を始めるので、1 focal plane から離れた位置に存在する物体部分にはぼけが生じてしまう。この問題を解決するために、LFR の画像合成を複数の focal plane において行い、それにより得られた画像を画素ごとに最も焦点が合っている画像を選択し、その画素を読み出すことによって全体的な画像を求める方法が提案されている。本稿では中間視点画像が正しい奥行きの前後においてぼけが対称であることを利用した新しい合焦判定尺度を提案し、その有効性をシミュレーションで評価する。

figure1 に LFR のアルゴリズムを示す。入力画像は等間隔に配置されたカメラによって取得され、画像合成はレンダリングカメラを任意の位置におき、その位置に到達する個々の光線の色を求めることによって実現される。この手法は空間全体を一様な平面 (focal plane) と仮定して扱うので、物体の三次元構造を用いることなく、簡単な処理で対応点の計算が可能である。しかし、LFR は focal plane 付近の物体は鮮鋭に合成されるが、focal plane から離れた物体の場合、対応点の不一致が生じ、画像に劣化が生じてしまう [2] [3] [4]。上記の問題を解決するための手法としてステレオマッチングと depth-from-focus [5]がある。ステレオマッチングは各画素の設定した仮想平面における対応点を計算し、マッチングの度合いが一番高いものを正解の奥行きとする方法である。この手法は全ての画素、全ての奥行きとの対応点を計算するため、検索が面倒な部分がある。depth-from-focus は LFR による画像合成を複数の focal plane において行い、それにより得られた画像を画素ごとに最も焦点が合っている画像を合焦判定（焦点があっているかどうかの判定）により選択し、その画像を読み出すことによって全体的に鮮鋭な画像を求める方法である [7] [8]。この手法の利点は合焦判定に対応点の計算が必要ないことである。しかし、問題となっているのは、どのような合焦判定を用いて焦点の合っている領域を検出するか、という点である。LFR による合成画像の場合、一般の画像で用いられる、高周波成分を利用した合焦判定 [6]が機能しない (figure 2)。これは、合成画像のぼけは高周波成分を持つ二重像（ゴースト）を含んでいるからである。

よって、本稿では LFR の中間視点画像生成に特化し、中間視点画像における LFR のぼけが正しい奥行きの前後において同じになること（ぼけの対称性）を利用した新しい合焦判定法を提案し、その判定法を用いた全焦点画像の品質の定量的評価を行った。

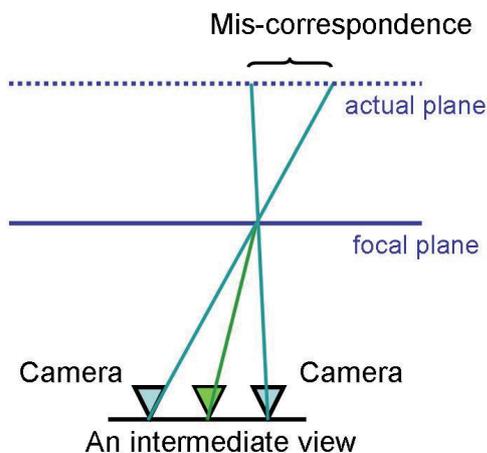


Figure 1 Principle of LFR



Figure 2 the synthesized image by using general focus measure

2 提案手法

2-1 合焦判定の原理

提案手法が LFR で得られた画像群をどのような合焦判定を行い、合焦距離を決定するか説明する。LFR で合成した画像群のある画素におけるぼけの様子を figure3 に示す。ここで、 z_l は focal plane の位置で

あり、物体が z_l の物体の奥行きにあるときを示している。focal plane が実際の物体のシーンの近くでは鮮鋭に再構成できるが、シーンから遠ざかるにつれて、非対応点を参照してしまいぼけが生じる。そのぼけ方が中間視点画像の場合、focal plane を適切に設定すると、正しい奥行きの前後で同じ劣化を示す。よって、

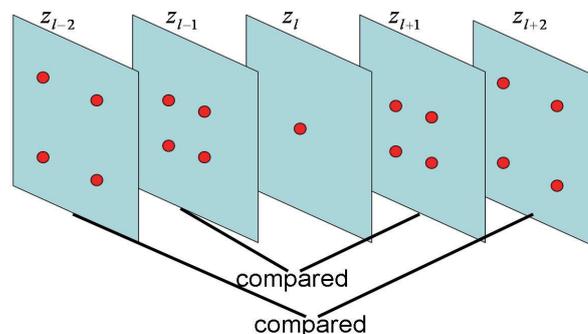


Figure 3 Shift-variant LFR images

ある奥行きに対してその前後の奥行きを比較し、その差分値を評価することで、焦点の合っている領域を検出することができる。提案手法では差分値をグラフにあらわしたときに最も前後の値と変化の大きい極小の値を持つ奥行きが合焦していると判断する。ここで、最小値でなく極小値を用いる理由は、正しい奥行きから大きく離れた奥行きでは再構成画像の変化が極小であるため、正解付近の奥行きにおける差分より小さくなるためである。

2-2 全焦点画像合成手順

1. まず、あらかじめ候補となる focal plane の奥行き $z_n (n = 0, 1, \dots, N-1)$ を設定する。 z_n は、視差空

間（奥行きの数）を等分割にするように、次式で設定する。このように配置することにより、中間視点画像において、前述したぼけの対称性を満たすことが確認されている。

$$\frac{1}{z_n} = \frac{1}{z_{\max}} + \left(\frac{1}{z_{\min}} - \frac{1}{z_{\max}} \right) \times \frac{n}{N-1} \quad (1)$$

Z_{\max} , Z_{\min} は、配置する focal plane の奥行き の最大値と最小値である。

2. それぞれにおいて LFR による中心からの画像合成を行う。ここで、 Z_n において合成された画像を $G_n(x, y)$ とあらわすことにする。 (x, y) は画素の位置を表す。

3. 評価関数を式(2) に示す。ここで、 W は比較するブロックである。 Z_n の前後の s 番目の奥行きまでの合成画像の差分を取り、その結果を足し合わせる。

$$E_0^{(n)}(x, y) = \sum_{i=1}^s \sum_{(x,y) \in W} |G_{n-i}(x, y) - G_{n+i}(x, y)| \quad (2)$$

4. 式(2) を n の関数とみたとき、極小になるような n' を求める。そのような n' の中で、最もその前後の値との変化が大きい奥行き $n_{x,y}$ を求めることにより、最も焦点があう奥行きとする。

$$E^{(n)}(x, y) = E_0^{(n'-1)}(x, y) + E_0^{(n'+1)}(x, y) - 2E_0^{(n)}(x, y) \quad (3)$$

$$n_{x,y} = \arg \max_{n'} \{E^{(n)}(x, y)\} \quad (4)$$

5. 下記の式に従い、全焦点画像 $G(x, y)$ を得る。

$$G(x, y) = G_{n_{x,y}}(x, y) \quad (5)$$

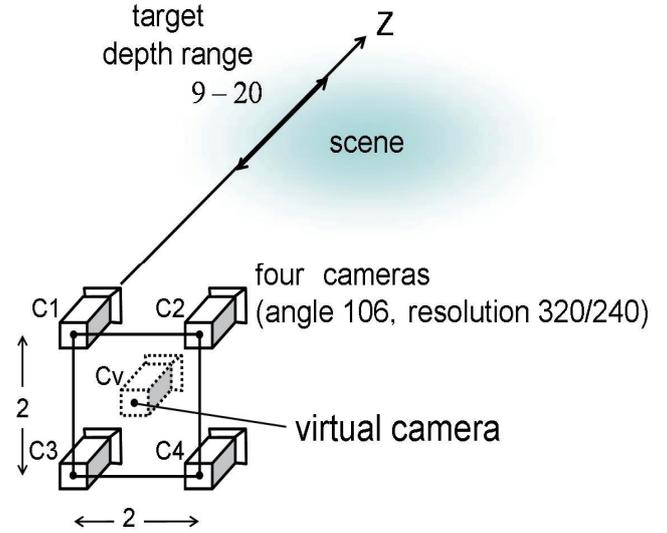


Figure 4 Configuration of the experience

3 シミュレーション

3-1 実験条件

実験の設定を figure4 に示す。ここで、本実験は実際の画像ではなく、CG で作った画像を用いており、対象シーンの存在する奥行き（カメラからの距離）は 9~20 とした。figure4 の位置関係において、カメラ間隔 2 で正方形に配置された 4 つのカメラの入力画像から得る。各入力画像は水平画角 106°、320×240 画素とする。配置する仮想平面は $N = 20$ とし、比較するブロック (W) のサイズを $3 \times 3, 5 \times 5, 7 \times 7, 9 \times 9, 11 \times 11$ 、比較する仮想平面の数を $s = 1, 2, 3, 4$ と変化させ、PSNR による画像品質の定量的評価を行った。ここで、focal plane の奥行きは式(1)を用いて設定し、 $Z_{\min} = 7$ 、 $Z_{\max} = 70$ とした。配置する focal plane を実際のシ

ーンの奥行きより広くとった理由は、前後の奥行きの差分によって評価をするため、実際のシーンより広い範囲の画像が必要になるためである。

3-2 比較する枚数、ブロックサイズによる影響

比較する focal plane の数(s)と、ブロックサイズ(BS)、合成画像の PSNR の関係を表した図を figure5 に示す。また再構成画像を figure8(a)~(d)に示す。BS、 s が大きいほど、PSNR の値は大きくなった。これは s , BS を大きくすると、参照箇所が増え、オクルージョン(対応点が存在しない部分)を含む割合が減少し、影響が低下することによって、画質が向上していると思われる。しかし、欠点として、 s を大きくすると、対象の奥行き前後に必要な画像の数が増えるため、候補となる奥行きの数が減少し、実際より広い範囲の奥行きをとる必要性が生じる。また BS を大きくすると、figure8(g) (h)のように、全体的に実際のシーンの奥行きよりも広がった奥行きを示した。しかし、PSNR は BS が大きいほど良い数値を示した。これは奥行きの広がった部分はそもそもオクルージョン部分でありどんな条件下でも正しく再構成できないため、画質に大きな影響を与えていないものと思われる。

3-3 極小値を用いる正当性

極小値を用いる正当性を示すため、提案手法と評価関数 $E_0^{(n)}(x, y)$ の最小値をとる n を最も焦点のあう奥行きとする手法と比較実験を行った。後者の手法による再構成画像を figure8(e)に示す。ここで BS は 5×5 、 $s = 4$ とした。Figure8(b)と(e)を比較すると、最小値を用いた画像は提案手法に比べエッジ部分に大きな劣化が生じており、提案手法の方が視覚的にも鮮鋭で PSNR 値も高かった。エッジ部分の画素における $E_0^{(n)}(x, y)$ のグラフを figure6 に、 $E^{(n)}(x, y)$ のグラフを figure7 に示す。この画素の正解の奥行きを検証したところ $n = 8$ であった。しかし最小値を用いた方法では、奥行き $n = 17$ が最小値なので、誤った部分を選択してしまっており、それが劣化の原因になっている。これは、正解の奥行きから大きく離れてしまうと、大きくぼけたもの同士を比較することになり、ぼけたもの同士の比較は、一般的に差分は小さくなる傾向があるため、正解の奥行きの前後の差分より小さくなってしまふことが原因である。それに比べ、式(3)を利用して評価した提案手法は、極小値で前後と最も値の変化が大きい奥行きが選択されるので、正解の奥行きである $n = 8$ が選ばれ、エッジ部分の劣化が改善されている。

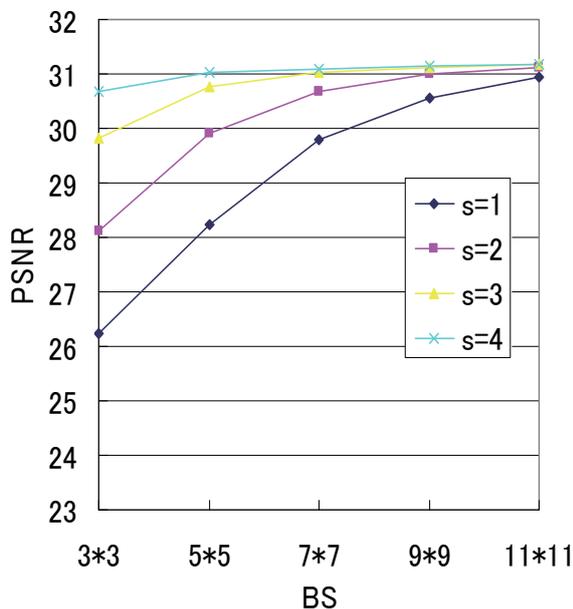


Figure 5 the relation of PSNR and the number of compared focus plane(s) and the size of block (BS)

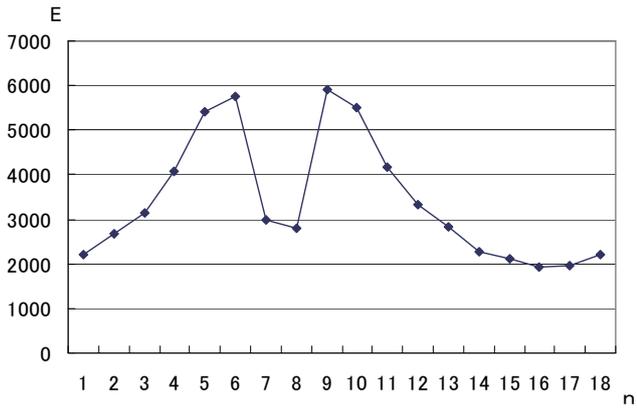


Figure 6 the evaluation function value $E^{(n)}(x,y)$ of a certain pixel

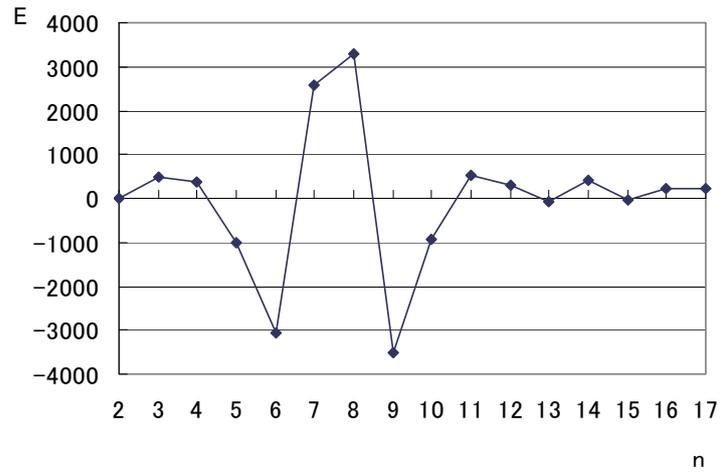


Figure 7 the evaluation function value $E^{(n)}(x,y)$ of a certain pixel



(a) BS 5*5 s 1



(b) BS 5*5 s 4



(c) BS 11*11 s 1



(d) BS 11*11 s 4



(e) BS 5*5 s 4
(the min method)



(f) Ground-truth



(g) BS 5*5 s 4
(depth map)



(h) BS 11*11 s 4
(depth map)

Figure 8 simulation results

4.終わりに

本研究では、多数のfocal planeの奥行きにおいてLFRによって合成した画像を用いて、ぼけ方の対称性を利用し全焦点画像を生成する方法を提案し、その有効性を確認した。[9]～[13]また、このぼけの対称性は、カメラ間の中心視点画像のみに成立する性質ではなく任意視点画像に拡張することも可能である。参照するブロックの位置を変えることで任意視点画像に対応する方法でも、その有効性を確認しているが、本稿においては記述を省略した[12]。

本研究より、目的としているような携帯端末でのモデル交換ベースでの簡易な物体認識機能の実現を可能とするためには、端末により取得した複数画像群より自由視点画像を再構成する際にLFR手法を援用することを想定すれば、交換すべき3次元対象物体モデルの情報表現制度はあまり高くする必要がない場合でも情報の再構成が可能であることを明らかにすることが出来た。

最後に、本要素技術の最終的な想定適用場面においては、粗い3D対象物体がモデル交換方式を前程として携帯電話端末のような比較的計算能力の高くない端末内部において、画像の認識等の処理を行わせるためのアルゴリズム間の負荷の最適化を計る必要がある。本実験で前程とした、物体のカメラからの相対的な存在位置の情報を、交換モデル情報から先験的に渡す方法、またそのようにして得られた情報を前程にして、LFRの計算処理を軽量化するための、s、BS等のパラメータの適応制御の効果等々についても、更なる研究が必要である。

【参考文献】

- [1] M. Levoy and P. Hanrahan. "Light Field Rendering", Proc. ACM SIGGRAPH, pp. 31-42, 1996.
- [2] A. Isaken, M. Leonard, and S. J. Gortler. "Dynamically Reparameterized Light Fields", Technical Report MIT-LOS-TR-778, 1999.
- [3] A. Isaken, M. Leonard, and S. J. Gortler. "Dynamically Reparameterized Light Fields", Proc. ACM SIGGRAPH, pp. 297-306, 2000.
- [4] J. Stewart, J. Yu, S. J. Gortler, and L. McMillan. "A New Reconstruction Filter for Undersampled Light Fields", Proc. EGSR, pp. 150-156, 2003.
- [5] K. Takahashi, A. Kubota, and T. Naemura. "A Focus Measure for Light Field Rendering", To appear in Proc. IEEE ICIP, Oct. 2004.
- [6] S.K. Nayer and Y. Nakagawa. "Shape from Focus", IEEE Trans. PAMI, 16, 8, pp. 824-831, 1994.
- [7] 高橋 桂太 久保田 彰 苗村 健 "Light field Renderingにおける合焦判定と全焦点画像合成", 電子情報通信学会誌 D-II Vol. J88-D-II No.3 pp.573-584, 2005
- [8] 高橋 桂太 苗村 健 "合焦判定に基づく全焦点 Light Field Viewer", FIT2004 J-010
- [9] 大澤 克俊, 久保田 彰, 羽鳥 好律, "中心視点画像生成のための合焦判定の一考察", 2006年度映像メディア処理シンポジウム (PCSJ/IMPS 2006), I2-12, pp.35-36, Nov. 2006.
- [10] 大澤 克俊, 久保田 彰, 羽鳥 好律, "合焦判定を用いた仮想視点画像の生成", 電子情報通信学会画像工学研究会 信学技法 IE2007-21, PP.1-6, June 2007
- [11] 大澤 克俊, 久保田 彰, 羽鳥 好律, "合焦判定を利用した中心視点画像の生成", 第6回情報科学技術フォーラム講演集, pp.253-254, Sep. 2007
- [12] 大澤 克俊, 久保田 彰, 羽鳥 好律, "合焦判定法を利用した仮想視点画像の生成", 2007年度映像メディア処理シンポジウム (PCSJ/IMPS 2007), I2-09, pp.41-42, Nov. 2007.
- [13] Oosawa Katsutoshi, Kubota Akira, Hatori Yoshinori, "GENERATING CENTER-VIEW IMAGE BY USING DEPTH-FROM-FOCUS", IEVC2007, 62.2P-3, Cairns, Australia, Nov. 2007

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
中心視点画像生成のための合焦判定の一考察	2006 年度映像メディア処理シンポジウム (PCSJ/IMPS 2006), I2-12, pp.35-36,	Nov. 2006.
合焦判定を用いた仮想視点画像の生成	電子情報通信学会 画像工学研究会 信学技法 IE2007-21, PP.1-6,	June 2007
合焦判定を利用した中心視点画像の生成	第 6 回情報科学技術フォーラム講演集, pp.253-254,	Sep. 2007
合焦判定法を利用した仮想視点画像の生成	2007 年度映像メディア処理シンポジウム (PCSJ/IMPS 2007), I2-09, pp.41-42,	Nov. 2007.
GENERATING CENTER-VIEW IMAGE BY USING DEPTH-FROM-FOCUS	IEVC2007, 62.2P-3, Cairns,	Nov.2007