

インテグラルイメージングにおけるピクセル再配置技術を用いた深度推定手法

井上耕太郎*・藤嶋 智晃*・Myungjin CHO**・李 旻 哲*

*九州工業大学院情報工学研究院電子情報工学研究系 〒 820-8502 福岡県飯塚市川津 680-4

**Department of Electrical, Electronic, and Control Engineering, IITC, Hankyong National University, 327 Chungang-ro, Anseong-si, Gyonggi-do 456-749, Republic of Korea

Depth Estimation of Integral Imaging Using Pixel Rearrangement Technique

Kotaro INOUE*, Tomoaki FUJISHIMA*, Myungjin CHO** and Min-Chul LEE*

*Department of Computer Science and Electronics, Kyushu Institute of Technology, Fukuoka 820-8502

**Department of Electrical, Electronic, and Control Engineering, IITC, Hankyong National University, 327 Chungang-ro, Anseong-si, Gyonggi-do 456-749, Republic of Korea

In this paper, we propose a new depth estimation technique using pixel rearrangement in integral imaging. In conventional depth estimation methods, the depth can be estimated by stereo matching with different perspective images of lenslet array based integral imaging. Since each perspective image has low resolution, the depth resolution of the estimated depth map will decrease. To solve this problem, we obtain the reconstructed 3D images with the same resolution as an image sensor by using pixel rearrangement technique. Therefore, we can estimate the high resolution depth map from multiple perspective images. For the depth estimation, the average derivative value which is the speed evaluation value of the laser speckle is used. It is a unique and effective approach to improve the depth resolution in integral imaging. Finally, to show the improvement of the depth resolution and accuracy in our proposed method, we implement the optical experiments.

Key words: integral imaging, depth estimation, three-dimensional reconstruction, three-dimensional recognition

1. はじめに

3D (three-dimensional) テレビやバーチャルリアリティーなどの普及に伴い、3D 映像技術に関する要求は日々高まりつつある。しかし、現在の 3D 技術の多くは眼鏡型のデバイスや、レーザーなどの特殊光源を必要とするものが多く、その利便性は高くない。そこで、裸眼での立体視が可能なインテグラルイメージング (integral imaging) が次世代の 3D 映像技術として着目され、盛んに研究されている¹⁻⁷⁾。

インテグラルイメージングは G. Lippmann によって発明された三次元空間像を記録・再生する手法である⁸⁾。それはレンズアレイ、イメージセンサー、そしてディスプレイのたった 3 つのデバイスで構成され、フルカラー、全視差情報、連続する視点をホログラフィーのようなレーザー光

源を用いることなく記録可能である。そのプロセスは、大きくピックアップと再構成の 2 つに分けられる。Fig. 1 にインテグラルイメージングシステムの概略図を示す。ピックアップでは、レンズアレイとイメージセンサーを用いて多視点画像を記録する。この画像は要素画像とよばれる。3D 空間像はディスプレイに要素画像を表示し、撮影に使用したレンズアレイを介することで得られる。これは光学的再構成 (optical reconstruction) とよばれ、光速で再構成画像を得られる大きな利点がある。しかし、大掛かりな光学系装置が必要であり、計算機上でのデータの応用性に乏しい。そこで、光学的再構成を計算機上でシミュレーションする計算的再構成 (computational reconstruction) が提案され、それを用いたさまざまな研究が進められている⁹⁻¹¹⁾。特に三次元空間における深度推定に関する研究

E-mail: lee@cse.kyutech.ac.jp

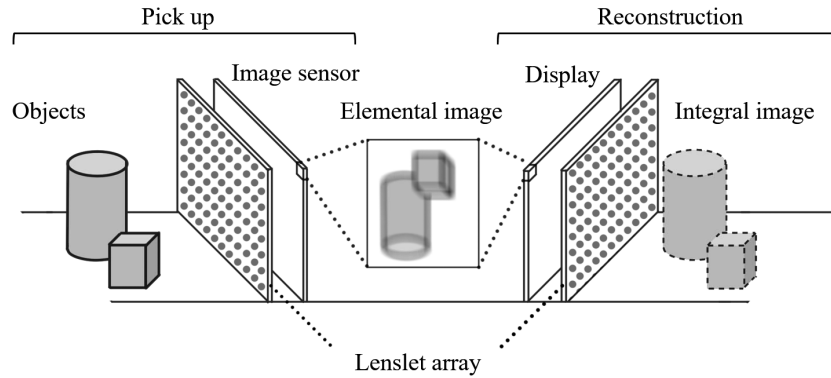


Fig. 1 Integral imaging system.

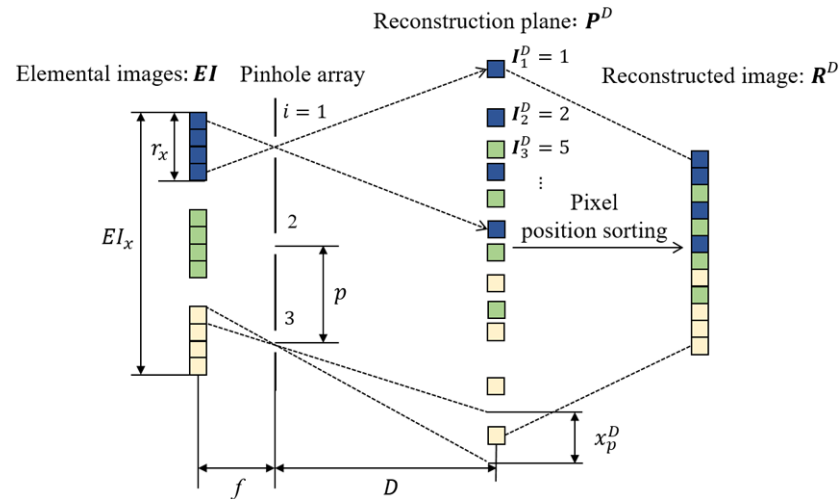


Fig. 2 Scheme of PERT.

は、遮蔽物除去やコンピュータビジョンといった大きな応用性が期待される重要課題である¹²⁻¹⁵⁾。

現在、インテグラルイメージングを用いた多くの深度推定手法は、VCR (volumetric computational reconstruction) とよばれる最も一般的な計算的再構成アルゴリズムを用いて研究されている¹⁶⁾。VCRは、再構成する距離に従って要素画像をすべて重ね合わせ、輝度の平均を計算することで再構成画像を得る。最もシンプルな計算方法であるが、すべての要素画像を重ね合わせるために、膨大な計算時間を必要とする。また、再構成画像の解像度は要素画像中の1視点の画像にほぼ等しいため、大幅に低下する。これらの問題を解消するため、PERT (pixels of elemental images rearrangement technique) が提案された¹⁷⁾。

PERTは要素画像を重ね合わせるのではなく、ピクセル単位で再配置することで再構成画像を得る手法である。重ね合わせや平均化処理が必要ないため、VCRと比較して非常に高速に再構成画像を得ることができ、再構成画像の解像度は要素画像と等しく高解像度である。再構成画像の

解像度が高いことは深度情報も高解像度であることを意味し、深度分解能についても大幅に改善される。そこで本研究では、PERTを用いた深度推定手法について提案し、その深度推定精度や解像度の改善を示す。

本論文の構成は以下の通りである。2章では提案する深度推定手法についての理論を述べる。3章では本提案手法の改善効果を確認するための検証方法について述べる。そして4章でその結果を示し、5章でまとめる。

2. PERTを用いた深度推定

2.1 PERT

PERTは要素画像のピクセル位置を再配置することで実現する。Fig. 2にその模式図を示す。ここでは簡単化のために、 x 軸のみの一次要素画像で考える。PERTは仮想的なピンホールアレイを介して、再構成距離 D におけるプロジェクション平面での要素画像間の実距離を計算する。ピンホールアレイによる拡大率 M^D はピンホールアレイまでの距離 f と再構成距離 D から、

$$M^D = D/f \quad (1)$$

となる。このとき、投影平面におけるピクセルの大きさ x_p^D は、 x 軸方向のセンサーサイズ C_x と要素画像の解像度 E_{I_x} から、

$$x_p^D = (M^D \times C_x) / E_{I_x} \quad (2)$$

である。以上から、投影平面での要素画像のピクセル位置行列 \mathbf{P}^D は

$$\mathbf{P}^D = n \times (x_p^D / 2) + (i-1) \times p, \text{ for } n = 1, 2, \dots, r_x \quad (3)$$

となる。ここで、 i は x 方向のピンホール数を、 r_x は要素画像の 1 視点の画像の解像度を、 p はピンホールのピッチを意味する。最後に \mathbf{P}^D はバブルソートを行い、再構成画像における要素画像のインデックス値を示す \mathbf{I}^D を得る。このインデックス値に従い、要素画像の再配置を行うことで、再構成画像 \mathbf{R}^D が得られる。本研究では他の再構成距離 D における再構成画像を順次得ることで、複数の焦点位置の再構成スライス画像を得る。この再構成スライス画像はそれぞれ焦点位置が異なるため、焦点が合う場所から深度情報を推定することができる。

PERT の最も大きな特徴として、要素画像と等しい解像度の再構成画像が得られる点がある。VCR では、再構成画像の解像度は 1 視点の要素画像の解像度に近くなるため、レンズ数が 10×10 のレンズアレイを用いた要素画像があった場合、イメージセンサーの解像度の 100 分の 1 という低解像度な再構成画像になる。しかし、PERT はその影響を受けず、高解像度な再構成画像を得ることができる。

2.2 AD 値による深度推定

本研究では、レーザースペックルの変動量評価指標のひとつである AD (average derivative) 値を再構成画像に適用する。AD 値は複数の撮影時刻にわたるある座標のピクセル輝度の変化量の総和を、輝度の平均値で除算することで得られる^{18,19)}。これはある時間範囲の輝度の明滅量を評価することに等しく、明滅の激しいピクセルほど AD 値は高くなる。本研究では、計算に用いる画像をスペックル画像から PERT で得られた再構成スライス画像とし、計算範囲基準を撮影時刻から深度へと変更した。Fig. 3 にその計算過程を、そして計算式を以下に示す。

$$AD_{(x,y)}^{D_n} = \frac{\frac{1}{N-1} \sum_{n=1}^{N-1} |L_{(x,y)}^{D_{n+1}} - L_{(x,y)}^{D_n}|}{\frac{1}{N} \sum_{n=0}^N L_{(x,y)}^{D_n}} \quad (4)$$

再構成距離 D における再構成画像の座標 (x, y) における輝度は $L_{(x,y)}^D$ とし、 N は計算に用いる深度範囲を示す。なお、実際に得られる AD 値の画像は微小な輝度変化がノイズとして現れるため、平均化フィルターにより平滑化する。

PERT で得られるスライス画像は非合焦位置ほど再配置

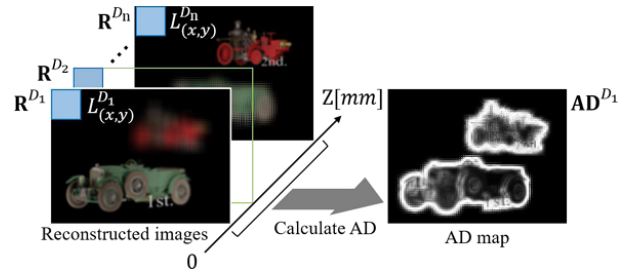


Fig. 3 Calculation process of AD.

における移動距離が大きいため、非合焦位置は明滅が激しく、逆に合焦位置は緩やかになる。つまり、再構成スライス画像に対して順次 AD 値を計算し、ある座標において AD 値が最小となるときに再構成距離が、深度推定結果 E^D として最も確かな値となる。これを式にまとめると、

$$E_{(x,y)}^D = \min (AD_{(x,y)}^{D_n}), \text{ for } n = 1, 2, \dots, N \quad (5)$$

となる。

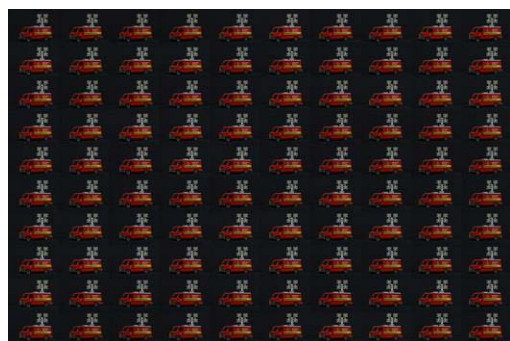
この手法の利点としては、再構成画像の 1 ピクセル単位での深度推定が可能であり、再構成画像はイメージセンサーと同等の解像度であることから、深度マップも同様に高解像度な情報が得られる点にある。最も一般的な深度推定手法にステレオ法が挙げられるが、ステレオ法は画像をある大きさのブロックに分割し、そのブロック単位での移動量を距離に置き換える。高解像度のステレオ画像が得られる場合は視差を精密に測定できるため問題ないが、要素画像がきわめて低解像度となるインテグラルイメージングにおいては、最適な手法ではないと考えられる。そこで、次章にて両者の比較を行う。

3. 検証方法

提案手法による深度推定の改善を示すため、従来手法によって得られた結果と比較する。従来手法とは要素画像に対してステレオ法によるマッチングを行い、そのピクセル移動量から深度推定をする手法である²⁰⁾。Fig. 4 にステレオ法による深度推定の概略図を示す。深度 d に存在するオブジェクトが、2つのピンホールによってイメージセンサーに結像したとする。 p はレンズアレイのピッチ、 δ はイメージセンサー上での結像位置間の距離、 f はピンホールの焦点距離を意味する。Fig. 4 から、底辺 δ と高さ $d+f$ による三角形と、底辺 p と高さ d による三角形の相似によって、オブジェクトまでの深度 d は

$$d = (p \times f) / (\delta - p) \quad (6)$$

である。ステレオ法の問題点は、深度方向の分解能が δ の分解能に依存する点である。要素画像が低解像度になるインテグラルイメージングでは、ステレオ法による深度推定



(a)



(b)

Fig. 4 Depth estimation by stereo method.

Table 1 Optical experimental setup.

Focal length of the camera	55 mm
Image sensor size	23.2×15.4 mm
Resolution of elemental image set	2100×1500 px
Pitch of camera array	1 mm
Number of elemental images	10×10
Distance between car and camera	350–410 mm
Distance between object and camera	210–230 mm

は多くの場合、十分な深度分解能と精度を提供できない。しかし、Jeon らによって提案されたライトフィールドカメラを用いた深度推定手法は、この問題を、位相シフト法を用いたサブピクセル単位のマッチングを行うことで解決している²¹⁾。彼らの提案した手法は 2015 年と比較的新しく、従来手法と比較して、きわめてなめらかな深度変化と高い精度を有する深度画像の推定に成功しており、プログラムも公開されていることから、その深度画像との比較を行う。

深度推定に用いる要素画像は光学実験によって得る。Table 1 に実験環境の設定を示す。オブジェクトは 1 台のミニカー（横 70 mm×縦 35 mm×奥行 30 mm）であり、遮蔽物のないクリアな視界と、松の枝による遮蔽物のある状況の 2 つを想定している。撮影は自動ステージ上に配置された一眼レフカメラによって行うが、本研究ではレンズアレイを用いて得られる要素画像を想定しており、1 視点

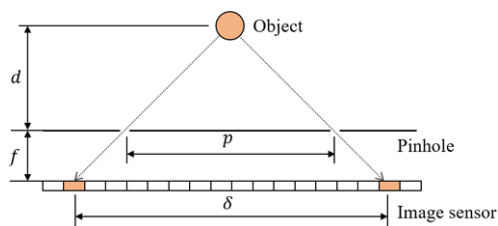


Fig. 5 Elemental images (Image size: 2100×1500 px). (a) without occlusion, (b) with occlusion.

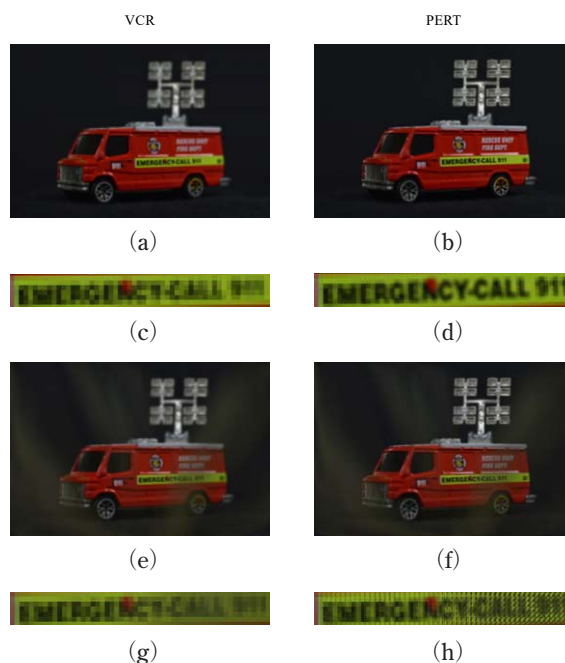


Fig. 6 Reconstructed images and enlarged images by VCR and PERT (Reconstruction depth: 400 mm). (a)-(d) without occlusion, (e)-(h) with occlusion.

あたりの画像の大きさは 210×150 px と低めにリサイズしている。Fig. 5 に光学実験で得られた要素画像を示す。検証は、この要素画像から推定された深度画像によって行う。

4. 検証結果

検証結果を得た際の提案手法と、従来手法のアルゴリズムにおける設定について述べる。提案手法は AD 値の計算深度範囲を 100 mm ごと、再構成画像間の深度間隔を 20 mm、平均化フィルターの大きさを 10 px 四方に設定した。使用した再構成画像の深度範囲は 300～500 mm である。従来手法は最適化に関する多くのパラメーターをもち、そのほとんどは Jeon らの論文²¹⁾ に従うが、われわれの光学実験に最適化するため、深度のラベル数とサブピクセル単位をそれぞれ 40, 0.06 とした。

提案手法では深度推定に再構成画像が必要であるため、Fig. 6 に再構成画像を示す。Fig. 6 において、(a) と (e) は

最も一般的な計算的再構成手法である VCR による再構成画像であり、(b) と (f) は提案手法で用いた PERT による再構成画像である。なお、VCR と PERT では再構成画像のアスペクト比が異なるため、同様の大きさの三次元像となるように、VCR 側の再構成画像はリサイズされていることに留意されたい。VCR によって得られる再構成画像の解像度は要素画像中の 1 視点におおよそ等しく、 228×158 px となる。一方、PERT による再構成画像の大きさは常に要素画像と等しく 2100×1500 px であり、PERT によって得られる再構成画像はより高解像度である。これは、Fig. 6 において再構成画像の拡大図 (c) と (d) を比較すると、(d) で “EMERGENCY-CALL 911” の文字が確認できるのに対して、(c) ではそれが確認できないことから明らかである。この結果は、PERT によって得られた再構成画像を解析することによって、要素画像の 1 視点の大きさではなく要素画像の大きさで深度推定が可能であり、より高い深度分解能で深度推定が可能であることを示す。

Fig. 7 に、深度推定によって得られた深度画像を示す。深度画像は撮影視点から近いほど暗く、遠いほど明るい輝度で表現されている。Fig. 7 において、(a) と (b) の Jeon らの手法を用いた結果では非常になめらかな深度変化のある深度画像を推定しているが、(a) では車体の構造がやや不鮮明で、かつ、車体上部の投光器が背景と明確に分離していない。(b) の遮蔽物のある視界では車体のおおよその深度変化を示すものの、全体にわたって深度の大きな推定誤差がみられる。一方、提案手法の深度画像である (c) と (d) では全体的に粒状の深度誤差をもつ離散的な深度推定となっているが、(c) の遮蔽物のない視界では車体の詳細な構造が見て取れ、特に投光機は背景と明確に分離しており、深度推定が正確に行われていることがわかる。解像度で比較すると、Jeon らの手法では 210×150 px、提案手法は 2100×1500 px と 100 倍の解像度を有し、本手法の利点を確認できる。(d) の遮蔽物のある視界では全体にわたって大きな誤差がみられ、車体後部は (c) で深度変化がみられたものの、(d) では損なわれている。

提案手法は遮蔽物のない視界において高解像度な深度画像を推定することができるが、遮蔽物のある状況や色変化の急激なエッジ部分において多くの深度がやや遠くに推定される問題がある。この原因は AD 値の特性にあると考えられる。遮蔽物がなければ PERT による合焦位置周辺の画素はおおよそ似た色と輝度のピクセルが配置されるはずだが、遮蔽物が存在することにより合焦位置周辺にもまれに非合焦位置のピクセルが配置され、AD 値を意図せず高めてしまう。この解決には、現在輝度だけで算出している

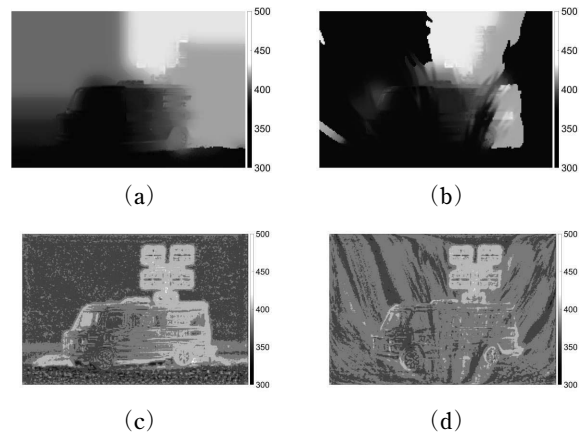


Fig. 7 Comparison of depth maps. (a)-(b) Jeon *et al.*, (c)-(d) Proposed method, (a)(c) without occlusion, (b)(d) with occlusion.

AD 値に対し、色を考慮する必要がある、将来解決すべき課題のひとつである。また、エッジ部分も同様の傾向があるが、こちらも同様に、AD 値の算出にあたり急激な色変化のある部分の AD 値の減少速度が遅くなるため、本来あるはずの深度位置から遠くに推定される。色情報を考慮することで低減される問題ではあるが、こちらは隣接ピクセルとの関係を考慮する確率場による最適化が最も効果的である。現在算出されている深度は、単一のピクセルの深度方向に対して算出された AD 値を最小化する位置であり、これは局所最適解である。AD 値を用いた大域最適解の手法の考案も次なる課題である。

5. ま と め

本論文では、インテグラルイメージングにおける新たな深度推定手法として、要素画像のピクセル再配置によって得られる再構成画像から AD 値を用いて深度推定する手法を提案した。本提案手法は、インテグラルイメージングにおいてイメージセンサーの解像度と同程度の深度画像を得ることが困難だった問題を解決するものであり、イメージセンサーと同じアスペクト比、かつ解像度で得られる深度画像の再利用性や応用性は、従来の手法をはるかに上回るものである。光学実験では、遮蔽物のない状況で良好な深度推定を示した。しかし、遮蔽物のある状況や色変化の急激なエッジ部分では、AD 値が減少するはずの合焦位置にもかかわらず AD 値が高く算出される問題があり、深度推定の精度を低下させる問題がみられた。今後の課題としては、AD 値を計算するため手作業で決定している深度範囲や深度間隔を自動で決定する手法や、色を考慮した AD 値の算出および大域最適化手法の考案が挙げられる。

文 献

- 1) B. Javidi, P.-D. Rodrigo and S.-H. Hong: *Opt. Lett.*, **31** (2006) 1106-1108.
- 2) M. Cho and B. Javidi: *Opt. Lett.*, **33** (2008) 2737-2739.
- 3) M. Cho and B. Javidi: *J. Disp. Technol.*, **5** (2009) 345-349.
- 4) J. Jung, M. Cho, D. K. Dey and B. Javidi: *Opt. Lett.*, **35** (2010) 1825-1827.
- 5) Y. Piao, M. Zhang and E.-S. Kim: *Optic. Laser. Eng.*, **50** (2012) 1602-1610.
- 6) M.-C. Lee, K. Inoue and M. Cho: *J. Inf. Commun. Convergence Eng.*, **13** (2015) 270-274.
- 7) M.-C. Lee, K. Inoue, N. Konishi and J.-J. Lee: *J. Inf. Commun. Convergence Eng.*, **13** (2015) 275-279.
- 8) G. Lippmann: *C. R. Acad. Sci.*, **146** (1908) 446-451.
- 9) S.-H. Hong and B. Javidi: *Opt. Express*, **12** (2004) 4579-4588.
- 10) D.-H. Shin and E.-S. Kim: *J. Opt. Soc. Korea*, **12** (2008) 131-135.
- 11) J.-J. Lee, B.-G. Lee and H. Yoo: *Appl. Opt.*, **50** (2011) 1889-1893.
- 12) C. Wu, A. Aggoun, M. McCormick and S.-Y. Kung: *Proc. SPIE*, **4660** (2002) 135-146.
- 13) J.-J. Lee, B.-G. Lee and H. Yoo: *Appl. Opt.*, **50** (2011) 5624-5629.
- 14) D.-H. Shin, B.-G. Lee and J.-J. Lee: *Opt. Express*, **16** (2008) 16294-16304.
- 15) J.-S. Lee, J.-H. Ko and E.-S. Kim: *Opt. Commun.*, **191** (2001) 191-202.
- 16) S.-H. Hong, J.-S. Jang and B. Javidi: *Opt. Express*, **12** (2004) 483-491.
- 17) M. Cho and B. Javidi: *J. Disp. Technol.*, **5** (2009) 61-65.
- 18) 藤居 仁 : *Med. Imag. Technol.*, **10** (1992) 395-402.
- 19) 藤居 仁, 小西直樹 : *応用物理*, **66** (1997) 476-480.
- 20) M. Z. Brown, D. Burschka and S. D. Hager: *Trans. Pattern Anal. Mach. Intell.*, **25** (2003) 993-1008.
- 21) H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai and I. S. Kweon: "Accurate depth map estimation from a lenslet light field camera," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Boston, 2015) pp. 1547-1555.