

多視点照度差画像を用いた光源方向推定

岩野 俊介^{1,a)} 小山田 雄仁¹ 望月 義彦¹ 石川 博¹

概要: 画像からの物体・シーンの3次元形状の復元において、代表的な手法である多視点ステレオ法と照度差ステレオ法では、物体に対するカメラの位置と光源環境の双方が異なるときには、正確な3次元形状を推定することができない。より正確な復元を行うためには、各画像ごとの物体に対する光源方向の推定が必要である。本研究では、多視点照度差画像から3次元形状を復元するために必要な光源方向を、入力画像のみを用いて推定する手法を提案し、合成画像を使った予備的な実験の結果を報告する。

1. はじめに

1.1 背景

コンピュータビジョンは、コンピュータによる視覚の実現を目的とする分野である。この分野において、画像から物体の3次元形状を復元する問題は主要な課題として研究されている。3次元形状の復元は技術と計算機性能の進歩に伴って高精度化・高速化され、研究の成果はコンピュータグラフィクス(CG)、拡張現実(AR)といった映像メディアを始め、ロボット工学、医療、教育など様々な分野にフィードバックされている。

3次元復元手法の一つである多視点ステレオ法では、対象となる物体やシーンを多数の視点から撮影した複数の画像から対象物体の3次元形状の復元を行う。現在では、画像のみから大規模な建造物や都市の形状復元が可能であり、インターネット上に存在する大量の画像から形状復元を行う研究も行われている [1]。

3次元復元の応用先のひとつである拡張現実とは、現実世界の映像にCGによってレンダリングした仮想物体を重畳表示することで現実世界の情報を拡張する。従来のCGでは物体の幾何形状、反射特性、環境の光源分布などが与えられた上でその物体の画像をレンダリングするが、これに対し、実物体や周辺環境の画像をもとにして物体の形状、反射特性、環境分布を推定していくインバースレンダリング [2] というシーン解析のアプローチがARの分野で注目されている。ARにおいて視覚的に現実感の高い画像を生成することは中心的な課題のひとつであり、現実世界と整合性の高い仮想物体のレンダリングを実時間で行うための研究がされてきた [3]。仮想物体とシーンの光学的整合性

表 1 多視点照度差ステレオ法

著者	光源位置	光源数	備考
Hernandez [4]	未知	1	
Zhou [5]	未知	1	同じ視点からの照度差画像を多視点分用いる
Higo [6]	既知	1	光源付き手持ちカメラ
Yoshiyasu [7]	既知	1	光源付き手持ちカメラ
Li [8]	既知	1	非ランバート反射物体
Wu [9]	既知	多数	
Fu [10]	既知	多数	運動している対象物体

の問題を解決するために、シーン中の光源環境の推定が必要である。

1.2 目的

本研究では、物体に対する光源方向を推定することを目的とする。3次元形状と光源方向の情報を組み合わせることで、実シーンへの光学的整合性の高いARオブジェクトの表示の他、対象物体の表面反射率が既知ならば復元したモデルのリライティングを行うことができる。

1.3 関連研究

光源情報を取り扱った多視点ステレオ法は多視点照度差ステレオ法 (Multi-View Photometric Stereo; MVPS) と呼ばれる。多視点照度差ステレオ法の研究内容には仮定する光源環境の違いがある。既存研究の相違点を表1にまとめる。

Hernandezらの手法 [4] は、本研究と同じく、方向が未知の単一平行光源下で撮影した画像を用いている。Hernandezらの手法では、まず Visual Hull から得られる対象物体の大まかな3次元形状と表面法線方向から光源方向を推定する。その後、推定した光源方向を用いて復元点の3次

¹ 早稲田大学大学院 基幹理工学研究科 情報理工・情報通信専攻
^{a)} s.iwano.162@toki.waseda.jp

元位置と表面法線方向について最適化を行い、それぞれの正確な推定値を求めている。

2. 理論

ここでは、画像からの3次元形状復元の代表的な手法である多視点ステレオ法 (Multi-View Stereo) と照度差ステレオ法 (Photometric Stereo) の理論について述べる。

2.1 多視点ステレオ法

多視点ステレオ法 (Multi-View Stereo) は、視点の変化に伴う対象物体の見えの変化を利用し、異なる視点から撮影された複数枚の画像から対象物体上の点の3次元位置を推定する手法である。光源環境は一定であると仮定し、複数枚の画像と、既知である撮影したカメラの内部・外部パラメータを用いて、画像上の点の3次元位置を推定することができる。

ある3次元点 \mathbf{X} が画像 \mathbf{I} 上の画素 \mathbf{u} として観測される時、 \mathbf{X} と \mathbf{u} の関係は、画像を撮影したカメラの内部パラメータ \mathbf{K} とカメラの視点を表す外部パラメータ (\mathbf{R}, \mathbf{t}) を用いて、

$$\mathbf{u} \cong \mathbf{K}[\mathbf{R}|\mathbf{t}]\mathbf{X} \quad (1)$$

と表せる。カメラの内部及び外部パラメータ及び各画像間における点の対応関係が分かっている時、ある点の3次元点位置 \mathbf{X} は各画像における画素位置 $\mathbf{u}_i, i = 1, \dots, N$ と式 (1) から

$$\hat{\mathbf{X}} = \underset{\mathbf{X}}{\operatorname{argmin}} \sum_{i=1}^N \|\mathbf{u}_i - \mathbf{K}[\mathbf{R}_i|\mathbf{t}_i]\mathbf{X}\|^2 \quad (2)$$

と推定できる。

2.2 照度差ステレオ法

照度差ステレオ法 (Photometric Stereo) は、光源環境の変化に伴う対象物体の見えの変化を利用し、異なる照明環境下で撮影された複数枚の画像から対象物体の表面形状を推定する手法である。照度差ステレオ法では物体表面の法線ベクトルを画素毎に計算するため、対象物体表面の3次元位置を推定する多視点ステレオ法に比べて物体表面の形状をより微細に推定することができる。その反面、照度差ステレオ法で推定するのは面の傾きであるため、物体の奥行きを直接求めることはできない。

照度差ステレオ法では、シーンとカメラを固定し、異なる単一平行光源下で撮影した複数枚の画像を用いる。

$$\mathbf{b} = \mathbf{s}^\top \cdot \rho \mathbf{n} \quad (3)$$

単一平行光源の位置を変えながらランバート反射する物体を N 回撮影する時、対象物体上のある点の画像上の輝度値 $b_i, i = 1, \dots, N$ は注目点の法線ベクトル \mathbf{n} 、アルベド ρ 、

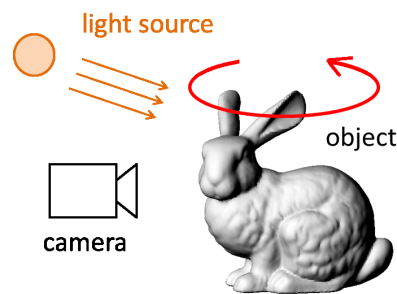


図 1 撮影シーンの概略図。運動する対象物体を固定されたカメラで撮影する。画像に用いたモデルは Stanford Computer Graphics Laboratory の”Stanford Bunny”[11].

光源方向 \mathbf{s} を用いて

$$\begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} = \begin{bmatrix} \mathbf{s}_1^\top \\ \vdots \\ \mathbf{s}_N^\top \end{bmatrix} \rho \mathbf{n} \quad (4)$$

と表せる。光源方向 \mathbf{s}_i が既知である時、注目点の法線ベクトル \mathbf{n} はスケールに関する曖昧性を残し、

$$\hat{\mathbf{m}} = \begin{bmatrix} \mathbf{s}_1^\top \\ \vdots \\ \mathbf{s}_N^\top \end{bmatrix}^+ \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} \quad (5)$$

と推定できる。ここで、 \mathbf{A}^+ は行列 \mathbf{A} の疑似逆行列を表す。法線ベクトルのノルムが1である性質を利用し、物体の法線ベクトルとアルベドは

$$\hat{\rho} = \|\hat{\mathbf{m}}\| \quad (6)$$

$$\hat{\mathbf{n}} = \frac{\hat{\mathbf{m}}}{\|\hat{\mathbf{m}}\|} \quad (7)$$

と推定できる。

3. 提案手法

本研究では、対象物体に対する光源環境の異なる多視点画像から、対象物体の3次元形状と対象シーンの光源方向を推定する手法を提案する。図1に対象とする撮影環境を示す。固定された単一平行光源によって照らされながら運動している物体を時々刻々と撮影している。対象物体の表面反射特性はランバート反射に従い表面反射率が一定であると仮定する。また、光源は単一の無限遠平行光源を仮定する。

手法の流れを図2に示す。上述した環境で撮影された画像 (図2a) を入力データとする。まずはじめに Structure-from-Motion (SfM) [12] によって、入力画像群からカメラの内部パラメータ \mathbf{K} と各画像間での対象物体の移動 $(\mathbf{R}_i, \mathbf{t}_i), i = 1, \dots, I$ を推定する。次に、カメラの内部パラメータ、SfMによって推定した $(\mathbf{R}_i, \mathbf{t}_i)$ 及び点の対応関係から多視点ステレオ (MVS) による3次元位置と法線ベクトル $(\mathbf{X}_j, \mathbf{n}_j), j = 1, \dots, J$ を推定する。最後に、SfM と

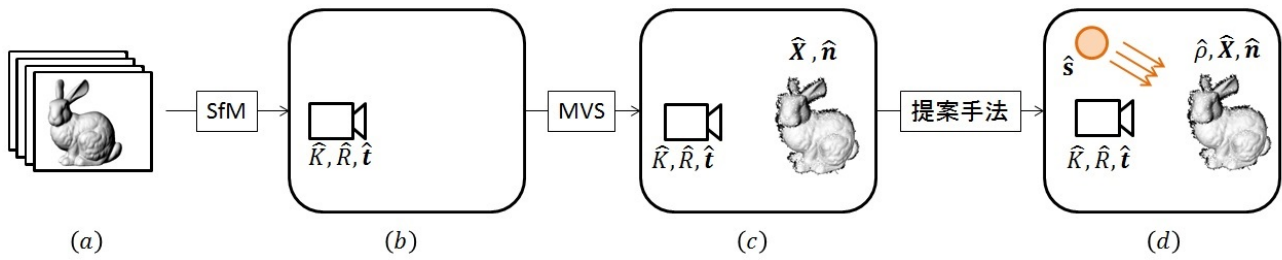


図 2 手法全体の流れの概要図. (a) 入力画像. (b) SfM によってカメラの内部・外部パラメータを推定する. (c) MVS によって物体の 3 次元形状を推定する. (d) 光源方向と表面反射率を推定する.

MVS によって推定された結果から光源方向 \mathbf{s} を推定する.

本論文は、この 3 段階の処理の内、最後の光源方向推定部を提案し、SfM と MVS に関しては既存手法を適用するものとする。まずはじめに 3.1 章で光源方向推定方法について述べ、次に 3.2 章で手法の実装について述べる。

3.1 光源方向の推定

提案する光源方向推定法は、物体上の点の 3 次元位置 \mathbf{X} 、表面法線ベクトル \mathbf{n} 、カメラの視線を表す外部パラメータ \mathbf{R} 及び入力画像 k 上で観測される輝度値 b を用いて、2.2 章で述べた照度差ステレオ法を基に光源方向を推定する。

対象物体の運動を考慮に入れると、対象物体上の点 \mathbf{X} は、

$$b = \rho(\mathbf{R}\mathbf{n})^\top \cdot \mathbf{s} \quad (8)$$

として観測されることになる。対象物体上の J 点が N 枚の画像上で観測される時、全観測値は列ベクトルとして

$$\begin{bmatrix} b_{1,1} \\ \vdots \\ b_{I,J} \end{bmatrix} = \rho \begin{bmatrix} (\mathbf{R}_1 \mathbf{n}_1)^\top \\ \vdots \\ (\mathbf{R}_I \mathbf{n}_J)^\top \end{bmatrix} \mathbf{s} \quad (9)$$

$$\rightarrow \mathbf{B} = \rho \mathbf{N} \mathbf{s} \quad (10)$$

と表せる。ここで、 \mathbf{B} は IJ 次元の観測行列、 \mathbf{N} は $IJ \times 3$ の法線方向行列を意味する。

光源方向 \mathbf{s} を求めるには、観測行列 \mathbf{B} に対し左から法線方向行列 \mathbf{N} の逆疑似行列をかける事でアルベドに関する曖昧性を残した光源方向 \mathbf{t} が

$$\hat{\mathbf{t}} = \mathbf{N}^+ \mathbf{B} \quad (11)$$

として求まる。光源方向とアルベドは、式 (5) と同様、 \mathbf{s} のノルムが 1 であることを利用し、

$$\hat{\rho} = \|\hat{\mathbf{t}}\| \quad (12)$$

$$\hat{\mathbf{s}} = \frac{\hat{\mathbf{t}}}{\|\hat{\mathbf{t}}\|} \quad (13)$$

と推定できる。

3.2 実装

ここでは入力画像から光源方向を推定する手法全体について述べる。

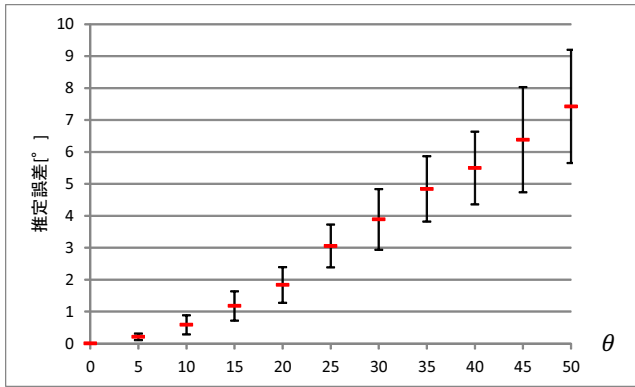
対象物体の 3 次元点位置と表面法線方向を得るために、Structure from Motion (SfM) ソフトウェアである Bundler [13] と、多視点ステレオ法のソフトウェアである Patch-based Multi-View Stereo version2 (PMVS2) [14] を用いる。Bundler は画像から抽出した SIFT 特徴量を用いてカメラの内部・外部パラメータを推定するソフトウェアである。PMVS2 は画像とそのカメラの内部・外部パラメータから物体の 3 次元形状を推定するソフトウェアである。PMVS2 は 3 次元位置と表面法線方向を持つパッチを用いて推定を行い、通常が多視点ステレオ法で推定される 3 次元位置に加えて、その点の表面法線方向も推定する。

4. 実験

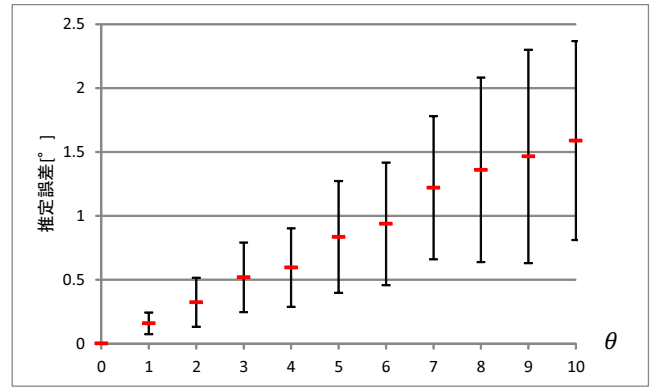
提案手法による光源方向の推定精度を確認するために、シミュレーション実験を行った。光源方向推定アルゴリズムの入力値である物体表面の法線方向、カメラ位置、及び画像の輝度値に対してノイズを加えた時に、提案手法によって推定される光源方向及びアルベドに与える影響を調べる。

画像の撮影回数 $I = 20$ 、対象物体上の観測点数 $J = 648$ とし、光源方向 \mathbf{s} は常にカメラに対して Y 軸回りに 45° の位置に設置する。各画像を撮影する時のカメラの位置を表す \mathbf{R}_i は、対象物体を中心とした円周上に均等に分布させ、物体の表面反射率 ρ は物体上で一様に 1 であるとし、各観測点の法線ベクトル \mathbf{n}_j は仰角と方位角がそれぞれ等間隔になるように生成した。輝度値 $b_{i,j}$ は、式 (3) から計算した。

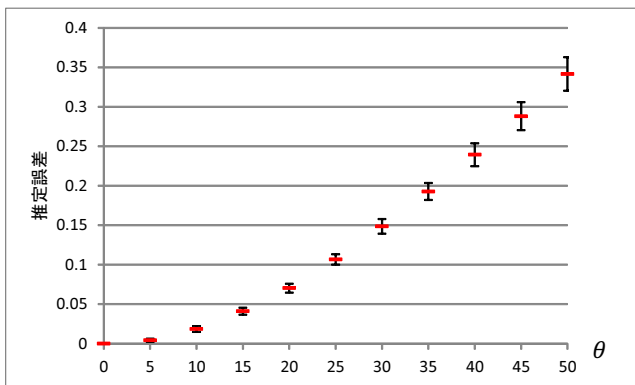
本実験では、法線方向 \mathbf{n}_j 、カメラ位置 \mathbf{R}_i 、輝度値 $b_{i,j}$ に対してそれぞれ独立にノイズを加えたデータに対して提案手法による光源推定を適用し、光源方向及び表面反射率の推定誤差による評価を行う。光源方向の推定値誤差は光源方向ベクトルの真値 \mathbf{s} と推定値 $\hat{\mathbf{s}}$ の成す角とし、表面反射



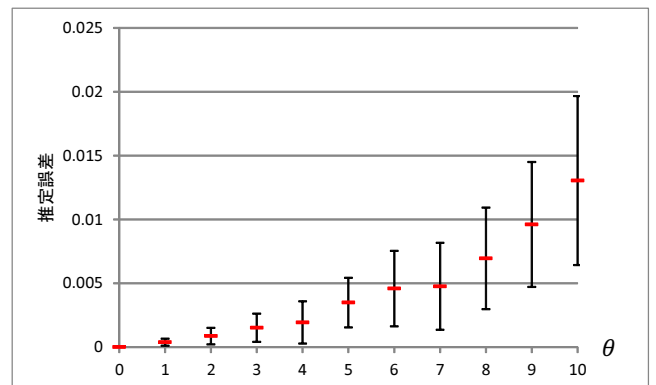
(a) 光源方向の推定誤差 (°).



(a) 光源方向の推定誤差 (°).



(b) 表面反射率の推定誤差.



(b) 表面反射率の推定誤差.

図 3 法線ベクトル上のノイズに対する光源推定誤差.

図 4 カメラ位置のノイズに対する光源推定誤差.

率の推定値誤差は真値 ρ と推定値 $\hat{\rho}$ の差の絶対値とする.

4.1 法線方向

ノイズを含んだ法線ベクトル \mathbf{n}'_j は法線方向ベクトル \mathbf{n}_j をランダムな方向 \mathbf{d} に θ° 回転させ生成した. ここで θ は平均 0, 標準偏差 θ の正規分布に従う. 光源方向, 表面反射率の推定結果を図 3 に示す.

光源方向の推定誤差は θ に比例して大きくなっている. 光源方向に比べ表面反射率の誤差は θ の影響が大きい. これは輝度値が光源方向ベクトルと法線ベクトルの内積に比例するため, 法線ベクトルのノイズの影響を大きく受けるからだと考えられる.

4.2 カメラの回転

ノイズを含んだカメラ位置 \mathbf{R}'_i はカメラ位置 \mathbf{R}_i にランダムな方向 \mathbf{d} に θ° 回転させる回転行列を掛けて生成した. ここで θ は平均 0, 標準偏差 θ の正規分布に従う. 光源方向, 表面反射率の推定結果を図 4 に示す.

法線にノイズを与えた場合に比べ, 推定誤差の標準偏差が大きい. これは設定したカメラ位置 \mathbf{R}_i の数が 20 で法線ベクトルの数 648 に比べて少ないことが理由として挙げら

れる. 式 (8) からわかるように, \mathbf{R}_i にノイズを与えると全ての法線ベクトル同じ方向にずれるため, 影響が大きいと考えられる.

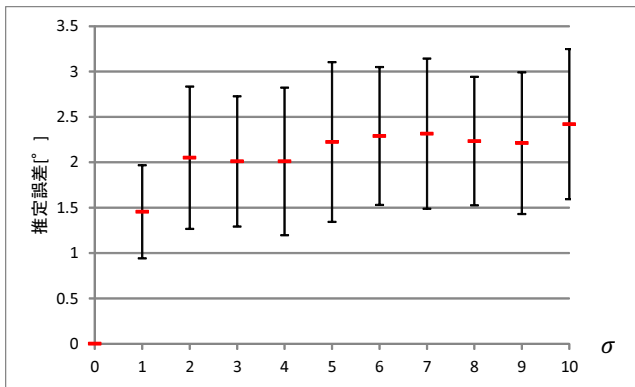
4.3 輝度値

ノイズを含んだ輝度値 $b'_{i,j}$ は輝度値 $b_{i,j}$ に σ を加算して生成した. ここで σ は平均 0, 標準偏差 σ の正規分布に従う. 輝度値は $[0, 1]$ の範囲にあるため, $b'_{i,j}$ が 0 以下または 1 以上になった場合は $b'_{i,j}$ をそれぞれ 0, 1 とした. 光源方向, 表面反射率の推定結果を図 5 に示す.

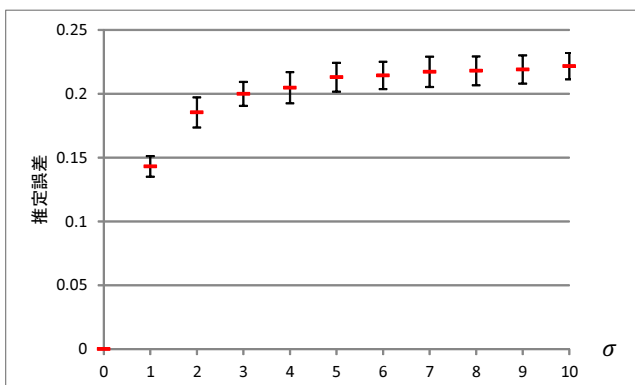
光源方向, 表面反射率の推定誤差が σ が大きくなるにつれて一定の値に収束しているが, これは $b'_{i,j}$ が $[0, 1]$ の範囲を超えた際に $b'_{i,j}$ をそれぞれ 0, 1 に設定したためだと考えられる. また, 輝度値が光源方向ベクトルと法線ベクトルの内積に比例するため, 輝度値に大きなノイズが含まれていても光源方向推定への影響が小さいのだと推察される.

5. おわりに

本論文では, 多視点画像から多視点ステレオ法で復元した 3 次元形状を用いる光源方向推定手法を提案した. 今後の課題として, 光源方向推定の高精度化, 光源方向を用い



(a) 光源方向の推定誤差 (°).



(b) 表面反射率の推定誤差.

図 5 輝度値のノイズに対する光源推定誤差.

た 3 次元形状の最適化があげられる.

参考文献

- [1] Furukawa, Y., Curless, B., Seitz, S. M. and Szeliski, R.: Towards internet-scale multi-view stereo, *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, IEEE, pp. 1434–1441 (2010).
- [2] 佐藤洋一, 向川康博: インバースレンダリング, 情報研報, Vol. 1106, pp. 2004–145 (2004).
- [3] 亀田, 蔵田, 清川, 稲見, 伴, 神原, 黒田, 前田: MR・AR・VR の現状 何ができて何ができていないのか? これから何をやるべきか?, 日本 VR 学会第 13 回大会論文集, pp. 461–465 (2008).
- [4] Hernandez, C., Vogiatzis, G. and Cipolla, R.: Multiview Photometric Stereo, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 30, No. 3, pp. 548–554 (2008).
- [5] Zhou, Z., Wu, Z. and Tan, P.: Multi-view Photometric Stereo with Spatially Varying Isotropic Materials.
- [6] Higo, T., Matsushita, Y., Joshi, N. and Ikeuchi, K.: A hand-held photometric stereo camera for 3-d modeling, *Computer Vision, 2009 IEEE 12th International Conference on*, IEEE, pp. 1234–1241 (2009).
- [7] Yoshiyasu, Y. and Yamazaki, N.: Topology-adaptive multi-view photometric stereo, *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, IEEE, pp. 1001–1008 (2011).
- [8] Li, G., Liu, Y. and Dai, Q.: Multi-view photometric stereo of non-Lambertian surface under general illuminations, *3D Imaging (IC3D), 2011 International Conference on*, IEEE, pp. 1–6 (2011).
- [9] Wu, C., Liu, Y., Ji, X. and Dai, Q.: Multi-view reconstruction under varying illumination conditions, *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, IEEE, pp. 930–933 (2009).
- [10] Vlasic, D., Peers, P., Baran, I., Debevec, P., Popović, J., Rusinkiewicz, S. and Matusik, W.: Dynamic shape capture using multi-view photometric stereo, *ACM Transactions on Graphics (TOG)*, Vol. 28, No. 5, ACM, p. 174 (2009).
- [11] Turk, G. and Levoy, M.: The Stanford Bunny (2005).
- [12] Snavely, N., Seitz, S. M. and Szeliski, R.: Modeling the world from internet photo collections, *International Journal of Computer Vision*, Vol. 80, No. 2, pp. 189–210 (2008).
- [13] Snavely, N.: Bundler-Structure from Motion software, *Web site:* <http://phototour.cs.washington.edu/bundler>.
- [14] Furukawa, Y. and Ponce, J.: Accurate, Dense, and Robust Multi-View Stereopsis, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 32, No. 8, pp. 1362–1376 (2010).