

インクリメンタル四面体カービング法による 三次元サーフェス生成

杉浦 貴行^{1,a)} 鳥居 秋彦^{1,b)} 奥富 正敏^{1,c)}

概要: Structure from Motion (SfM) 手法により, 単眼カメラ画像からその画像を撮影したカメラの位置姿勢と3D点の推定が可能である. SfMによって得られる点群から平面やテクスチャ情報を付加できるサーフェスを生成することは, 点群では表現しがたい遮蔽関係や変化の検出など, 様々な応用上重要である. 本論文では, 特に逐次的な入力に対して, 単純に毎フレーム始めから処理を繰り返すようなサーフェス生成ではなく, サーフェスの一部が拡張・更新されるインクリメンタルサーフェス生成の手法を提案する. 提案手法の特徴は, インクリメンタル処理による効率化を図りながら, 全体最適性の保たれたサーフェスを生成できる点にある.

1. はじめに

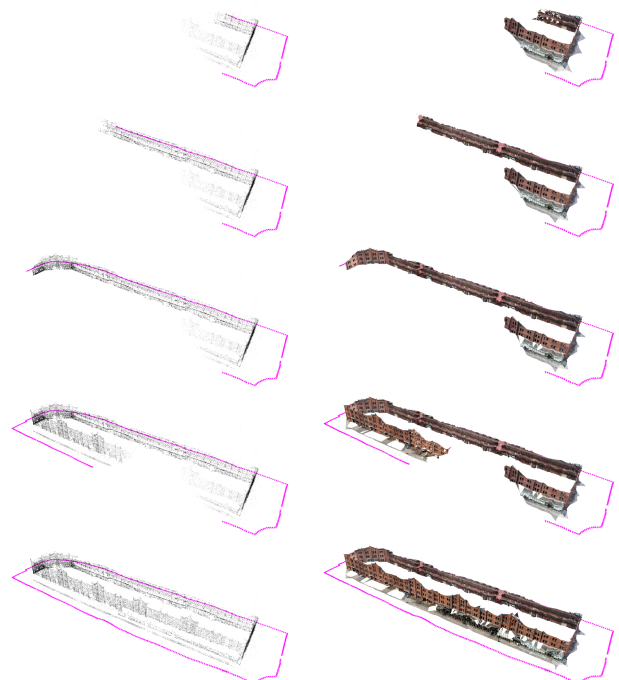
コンピュータビジョンの分野において, 多視点の画像を用いた三次元形状復元は盛んに研究されている. Structure from Motion (SfM) 手法 [1], [4], [14] により, 単眼カメラ画像からその画像を撮影したカメラの位置姿勢と三次元点を推定することができる. さらに三次元モデルとして, 点群では表現できない遮蔽関係やテクスチャ情報を付加できるサーフェスを生成する手法 [9], [10], [12], [17] も提案されている. こうした三次元モデルの復元は, AR [13] や大規模モデリング [1], [4], [5] など様々な応用上重要である.

特に連続した画像入力から逐次的に三次元モデルを生成しすることは, 変化検出 [15] や車両・UAVのナビゲーション [7], [19] などのアプリケーションに応用可能である. カメラ位置姿勢と三次元点の推定については, インクリメンタルに SfM を行う手法 [8], [14] が提案されている. サーフェス生成については, 個々に求めた基準フレームからのデプスマップを統合する手法 [13], [18] があるものの, これらの手法はシーン全体の一貫性を保っていない.

一方でマルチビューステレオによりシーン全体の一貫性を保ち, 三次元モデルの精度や完全性を追求する手法が存在する. 例えば, 平面走査法や疎な点群の拡張による点群の高密度化 [6], [17] や三次元ドロネー分割を利用したサーフェス抽出 [12], photo-consistency や幾何情報を利用した



(a)



(b)

(c)

図1 提案手法によるインクリメンタルサーフェス生成の様子. (a) 入力画像例. (b) SfMによって得られたカメラ位置姿勢と三次元点群. (c) 逐次的に生成されるサーフェス.

¹ 東京工業大学
Tokyo Institute of Technology

a) tsugiura@ok.ctrl.titech.ac.jp

b) torii@ctrl.titech.ac.jp

c) mxo@ctrl.titech.ac.jp

サーフェスのリファイン [9], [17] などがある. これらの手法で逐次入力に対してもシーンの一貫性を保つためには全

体を再計算する必要があり、インクリメンタル処理への拡張は困難な課題である。

本論文では、三次元ドロネー分割を利用した手法を拡張し、カメラ位置姿勢と三次元点群が逐次的に入力されると部分的にサーフェスを更新・拡張するインクリメンタルサーフェス生成の手法を提案する。提案手法では逐次的な処理による効率化を図っているにもかかわらず、コスト関数に対する全体最適解となるサーフェスを生成する。これを実現するために本論文では以下の二点を導入する。

- 視線と四面体の交差を検出する二つの方法を適応的に使用する。
- 孤立ノードの導入によりノードの増減に対応させたダイナミックグラフカット [11] をインクリメンタルサーフェス生成に適用させる。

提案手法の入力となるカメラ位置姿勢や三次元点群の推定にはインクリメンタル SfM を用いる。本論文ではサーフェス生成に焦点を当てるため SfM とサーフェス生成の統合は考えない。SfM は画像が追加されるたびにそのカメラ位置姿勢と三次元点群を追加復元し、インクリメンタルサーフェス生成の入力とする。SfM では既に復元されている点の消滅や融合をしないものとする。

2. 関連研究

マルチビューステレオで広く使われている手法の一つに PMVS [6] が挙げられる。この手法では疎な三次元点群のうち信頼度の高い点の周囲に新たに点を拡張することで、より密な三次元点群を生成する。Poisson サーフェス生成 [10] のような追加処理を行うことでサーフェスを得ることができる。また、近年注目を集めている手法に三次元点群に対する三次元ドロネー分割と物体の外側に属する四面体の削り出しによるサーフェス生成法が存在する [9], [17]。本論文ではこれを四面体カービング法と呼ぶ。四面体カービング法は対象の形状に対する事前知識を必要とすることなく複雑で大規模なシーンの復元が可能である。これらの手法は三次元モデルの精度や完全性に注視しているため、インクリメンタル処理への拡張は容易ではない。

一方で逐次的に三次元復元を行う手法もある。DTAM [13] はリアルタイムのカメラトラッキングと基準カメラにおけるピクセル単位のデプスマップ推定を行う。変化の少ない連続した入力画像が必要で、デスクスケールの復元のみ適用可能である。Hoppe ら [7] はインクリメンタル SfM と四面体カービング法を用いてインクリメンタル三次元復元を行い、サーフェスの精度による撮影者へのフィードバックを提案している。大規模シーンへの適用が可能であるが、サーフェス生成は毎フレーム始めから行うため、モデルの拡張による計算時間の増加は避けられない。Yu ら [19] も同様に四面体カービング法 (彼らは free

space carving と呼んでいる) を用いたインクリメンタル三次元復元を提案している。この手法ではモデルの拡張に対して一定の計算時間を実現しているが、四面体との交差の計算を行う視線をしきい値で制限しているため全体最適性が損なわれる。

本論文では、大規模シーンに対しても適用可能であり、全体最適性を保証しながら逐次的にサーフェスを生成する手法を提案する。

3. 四面体カービング法

四面体カービング法の概要について述べる (図 2)。まず三次元点群 $\mathcal{P} = \{p\}$ 、カメラ位置姿勢 $\mathcal{C} = \{c\}$ 、視線 $\mathcal{R} = \{r\}$ が入力される (a)。ここで視線 r とは、 p とその点を見ている c の位置を結んだ線分である (c)。 \mathcal{P} に対する三次元ドロネー分割によって空間を四面体群 $\mathcal{T} = \{t\}$ に分割する (b)。各四面体が対象物体の内部・外部どちらに属するのかが推定し、二つに分けられた四面体群の境界をサーフェスとして抽出する (d)。内外の推定は \mathcal{T} と \mathcal{R} の交差 (c) によって決まり、以下のコスト関数を最小にするバイナリラベリング問題としてとらえることができる：

$$E(\mathcal{L}) = \sum_{t \in \mathcal{T}} \{U_t(l_t) + \sum_{v \in \mathcal{V}_t} B_{t,v}(l_t, l_v)\}, \quad (1)$$

ここで、 $l_t \in \mathcal{L}$ は四面体 $t \in \mathcal{T}$ のラベルで、物体の内側 ($l_t = \text{IN}$) か外側 ($l_t = \text{OUT}$) のどちらにあるかを示す。 U_t は四面体に付けられるラベルの不一致に関するコストで、

$$U_t(l_t) = \begin{cases} N_{\text{in}}(t) & \text{if } l_t \neq \text{IN} \\ N_{\text{out}}(t) & \text{if } l_t \neq \text{OUT} \end{cases} \quad (2)$$

で表される。 $N_{\text{in}}(t)$ は延長線が t と交差する視線 r の数、 $N_{\text{out}}(t)$ は t が含むカメラからの視線 r の数をそれぞれ表す。この項は各四面体に視線との関係と矛盾がないラベルを付けるための項である。 $B_{t,u}$ は隣接する二つの四面体間のラベルの違いに関するコストで、

$$B_{t,v}(l_t, l_v) = \begin{cases} N_{\text{intrsect}}(t, v) & \text{if } l_t = \text{OUT} \cap l_v = \text{IN} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

で表される。 $N_{\text{intrsect}}(t, v)$ は t, v 間の面と t 側から v 側に向かう視線 r との交差の数である。この項は平滑化項として、隣接する四面体のラベルの変化を抑える。詳細は [17] を参照されたい。

このコスト関数に対して $s-t$ グラフカット [2] による全体最適化を行う。四面体群 \mathcal{T} をノードとするグラフを作成する。ノード間のエッジには隣接する四面体間の面に関する $B_{t,u}$ による重みを加え、ターミナル・ノード間には U_t による重みを加える。このグラフに対するグラフカットにより、コスト関数 E に関して全体最適解となるサーフェスを抽出することができる。

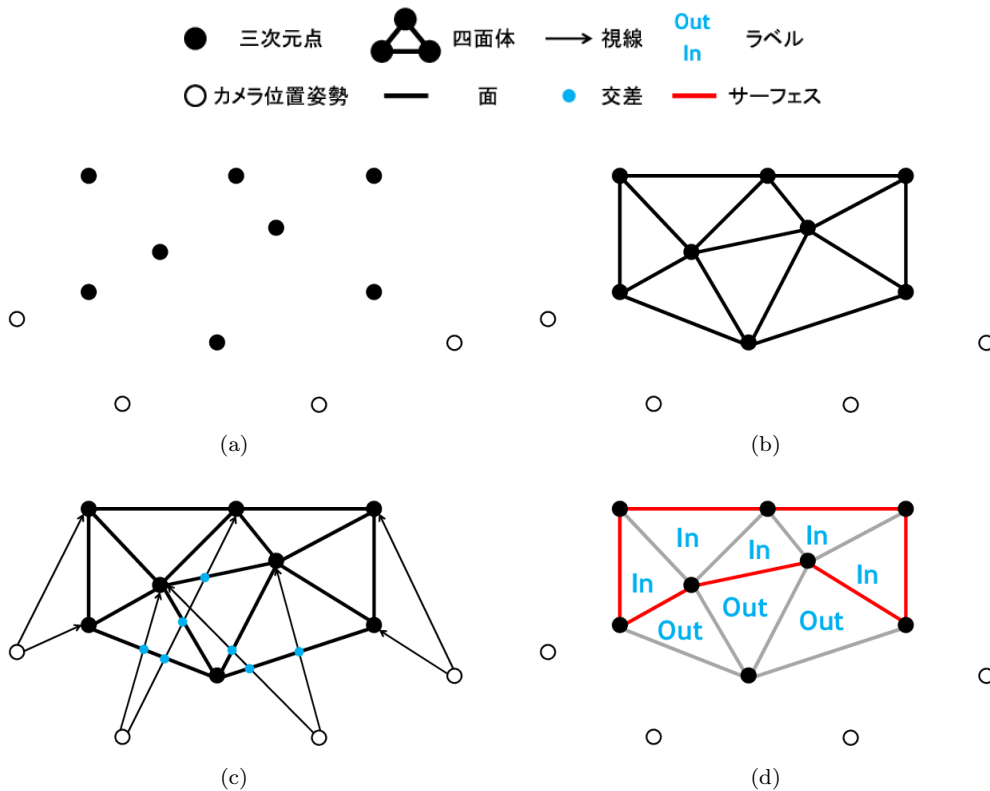


図 2 四面体カービング法の概要. (a) カメラ位置姿勢, 三次元群の入力. (b) 三次元ドロネー分割による空間の分割. (c) 四面体の面と視線の交差検出. (d) バイナリラベリングによる最適化及びサーフェス抽出.

4. インクリメンタル四面体カービング法

本章では, 四面体カービング法をインクリメンタル処理に拡張する方法を述べる. 提案手法では逐次的な入力に対して局所的な計算だけを行い, サーフェスの拡張・更新を実現する. その際, 抽出されたサーフェスが毎フレーム始めから処理する方法と完全に一致し, 全体最適性を保つようにアルゴリズムの拡張を行う. まず SfM で新たに推定されたカメラ \hat{c} と三次元点群 \hat{P} が入力されると, $\mathcal{P} \cup \hat{P}$ に対する三次元ドロネー分割により四面体群が更新される. \hat{c}, \hat{P} の追加と四面体群の更新により新たに発生した視線と四面体の交差を検出しグラフの重みの更新を行う. 最後に, 更新されたグラフに対してダイナミックグラフカット [11] を適用し全体最適なラベリングを求める. 以下, 提案手法において重要な二つの処理について詳しく述べる.

4.1 新しい交差の効率的な検出

四面体カービング法の中でボトルネックとなるのはグラフの重みを決めるための視線と四面体の面の交差検出である. インクリメンタル処理への拡張において重要なことは, 新たな視線と面の交差を検出しその部分に対応するグラフの重みだけを更新することである. そこで, 視線を二種類 $\mathcal{R}, \hat{\mathcal{R}}$ に分類する. 視線 $r \in \mathcal{R}$ は $c \in \mathcal{C}$ と $p \in \mathcal{P}$ を結ぶ線分である. $\hat{r} \in \hat{\mathcal{R}}$ は新たなカメラと三次元点が追加さ

れたことによる新たな視線である. 同様に四面体の面についても分類 $\mathcal{F}, \hat{\mathcal{F}}$ を行う. $f \in \mathcal{F}$ は既に存在していた四面体 \mathcal{T} の面で, $\hat{f} \in \hat{\mathcal{F}}$ は新たに生成された四面体 $\hat{\mathcal{T}}$ の面である. 提案手法では全体を再計算することなく, 新たな入力による変化を部分的に計算するため, すでに検出されている \mathcal{R} と \mathcal{F} の交差の再計算を避ける必要がある.

[9], [17] では, 交差検出として一本の視線が貫く四面体を検出する方法を用いている. この方法を本論文では “ray-wise tracing” と呼ぶ. p と c からなる視線 r について p を頂点に持つ四面体から c を内部に含む四面体まで, r と交差する面を持つ四面体をたどることで効率的に交差の検出が可能である. 検出するすべての交差が新しい交差となる $\hat{\mathcal{R}}$ には ray-wise tracing を適用する. 全体最適性を保証するために残りの \mathcal{R} と $\hat{\mathcal{F}}$ の交差を検出する必要があるが, ray-wise tracing は面の種類にかかわらず交差を検出するため, \mathcal{R} 全体に対して ray-wise tracing を適用することになり, 始めから行う手法と同じ計算量になる.

そこで, 一枚の面に対してどの視線が交差しているかを検出する方法を導入する. 本論文ではこれを “facet-wise detection” と呼ぶ. カメラの視野や三角面の visual cone を利用した視線の選択によって効率的に交差を検出する. この手法を $\hat{\mathcal{F}}$ に適用することで, 残りの \mathcal{R} と $\hat{\mathcal{F}}$ の交差を検出することができる.

まとめると, 新しい視線 $\hat{\mathcal{R}}$ に対して ray-wise tracing を,

新しい面 $\hat{\mathcal{F}}$ に対して facet-wise detection をそれぞれ適用することで \mathcal{R} と \mathcal{F} の交差を重複して計算することなく、全体最適性の保証に必要な新たな交差をすべて検出することができる。

4.2 ダイナミックバイナリラベリング

四面体群の更新と新たな交差の検出によりグラフが更新される。このグラフの更新はダイナミックグラフととらえることができ、ダイナミックグラフカット [11] によって局所的な流量の計算だけで最小切断を求めることができる。

提案手法ではエッジの重みの変化だけでなく、四面体の増減つまりノードの増減が起こる。これに対応するために孤立ノードを導入する。孤立ノードは接続されたすべてのエッジの重みを 0 としたノードである。ノードが消失したときはグラフから取り除く代わりに孤立ノードとし、新たに出現した時は孤立ノードのエッジに重みを加える。孤立ノードの導入によってダイナミックグラフカット [11] をグラフの構造の変化にも対応させることができる。

5. 実験

実験では、毎フレーム始めから処理を行うベースライン手法に対する提案手法の有用性を示す。また、他のマルチビューステレオ手法との定性的な比較を行う。

5.1 実験環境

実験は Intel Core i7, CPU 3.33GHz, RAM 24GB デスクトップ PC 上の Linux 64bit OS, C++で行った。実装において、OpenCV [3], CGAL [16], ダイナミックグラフカット [11] の各種ライブラリを利用した。

サーフェスの可視化において、提案手法によって抽出されたサーフェスに対して大きなパッチの除去と平滑化の二つの追加処理を加える。これらの処理はまだインクリメンタルに拡張されていないため、計算時間の評価からは除く。

実験で用いたデータセットは、焦点距離固定の単眼カメラから撮影された 978 枚の画像 (2296 × 1528 ピクセル) である。提案手法の純粋な評価を行うため、カメラ位置姿勢 \mathcal{C} と三次元点 \mathcal{P} はあらかじめ求めておく。そこから一台のカメラが入力されるとそれに関連した三次元点も入力される逐次的な入力を提案手法に加える。これは、SfM の全体バンドル調整における歪み除去や三次元点の融合・除去の影響を取り除くためである。

5.2 提案手法の評価

図 1 に提案手法によるインクリメンタルサーフェス生成の様子、図 4 に提案手法によるサーフェスを示す。逐次的な入力に対して目的とするインクリメンタルサーフェス生成を実現した。図 3 に提案手法 (黒) と毎フレーム始めか

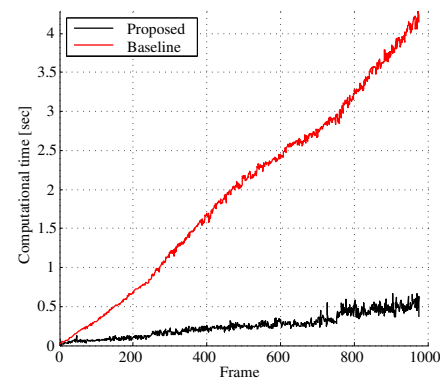


図 3 提案手法 (黒) と毎フレーム始めから処理を行うベースライン手法 (赤) の各フレームにおける計算時間。

ら処理を行うベースライン手法 (赤) による各フレームでサーフェス抽出にかかった計算時間を示す。画像の増加に対して計算時間が増えるベースライン手法と比較して、提案手法は計算時間の増加を抑えることができている。997 回の更新にかかった積算の計算時間は、提案手法が 245.82 秒、ベースライン手法が 1930.06 秒である。さらに、提案手法によって最終的に生成されたサーフェスがベースライン手法と完全に一致していることを確認した。

5.3 三次元モデルの定性的評価

提案手法と既存手法によって復元された三次元モデルを比較する。既存手法には公開されている PMVS [6] と CMPMVS [9] を利用した。各手法にカメラ位置姿勢と三次元点を入力した結果を図 5 に示す。CMPMVS については、[9] で提案されているカメラ位置姿勢を入力として平面走査法を行う手法 (d) と、提案手法と同じ入力になる平面走査法を行わない手法*1 (c) を記載した。提案手法による三次元モデルは他手法に劣らない結果を得ることができている。

6. おわりに

本論文では四面体カービング法を利用したインクリメンタルサーフェス生成の手法を提案した。視線と面の二つの交差検出法の適用的な利用と孤立ノードを導入したダイナミックグラフカット [11] の適用によって、全体最適性を保ちながら毎フレーム始めから処理を行う手法より十分高速にサーフェスを生成することができた。提案手法は既に復元された三次元点やカメラの変化に対応していない。今後の課題として、SfM のバンドル調整に対応しインクリメンタル三次元復元システム全体の統合が挙げられる。

謝辞 本研究は科研費 21240015 の一部助成を受けたものである。比較実験用プログラムを提供いただいた M. Jancosek 氏に感謝します。

*1 平面走査法を行わない CMPMVS は [9] の著者らの実装による。

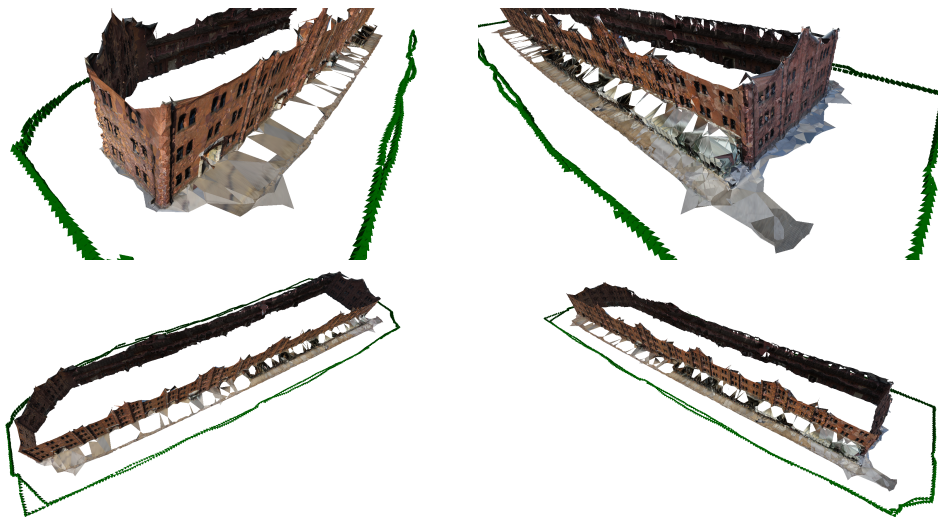


図 4 提案手法によるサーフェス. 緑の四角錐はカメラの位置姿勢を表す.



図 5 三次元モデルの定性的評価. (a) 提案手法. (b) PMVS[6]. (c) 平面走査法を行わない CMPMVS. (d) CMPMVS[9].

参考文献

- [1] Agarwal, S., Snavely, N., Simon, I., Seitz, S. and Szeliski, R.: Building Rome in a Day, *ICCV*, pp. 72–79 (2009).
- [2] Boykov, Y. and Kolmogorov, V.: An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision, *PAMI*, Vol. 26, No. 9, pp. 1124–1137 (2004).
- [3] Bradski, G.: The OpenCV Library, *Dr. Dobb's Journal of Software Tools* (2000).
- [4] Frahm, J.-M., Fite-Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.-H., Dunn, E., Clipp, B., Lazebnik, S. and Pollefeys, M.: Building Rome on a Cloudless Day, *ECCV*, pp. 368–381 (2010).
- [5] Furukawa, Y., Curless, B., Seitz, S. M. and Szeliski, R.: Towards Internet-scale multi-view stereo, *CVPR*, pp. 1434–1441 (2010).
- [6] Furukawa, Y. and Ponce, J.: Accurate, Dense, and Robust Multiview Stereopsis, *PAMI*, Vol. 32, pp. 1362–1376 (2010).
- [7] Hoppe, C., Klopschitz, M., Rumpel, M., Wendel, A., Kluckner, S., Bischof, H. and Reitmayr, G.: Online Feedback for Structure-from-Motion Image Acquisition, *BMVC*, pp. 70.1–70.12 (2012).
- [8] Indelman, V., Roberts, R., Beall, C. and Dellaert, F.: Incremental Light Bundle Adjustment, *BMVC* (2012).
- [9] Jancosek, M. and Pajdla, T.: Multi-view reconstruction preserving weakly-supported surfaces, *CVPR*, pp. 3121–3128 (2011).
- [10] Kazhdan, M., Bolitho, M. and Hoppe, H.: Poisson surface reconstruction, *Eurographics*, pp. 61–70 (2006).
- [11] Kohli, P. and Torr, P. H. S.: Efficiently Solving Dynamic Markov Random Fields Using Graph Cuts, *ICCV*, pp. 922–929 (2005).
- [12] Labatut, P., Pons, J.-P. and Keriven, R.: Efficient Multi-View Reconstruction of Large-Scale Scenes using Interest Points, Delaunay Triangulation and Graph Cuts, *ICCV*, pp. 1–8 (2007).
- [13] Newcombe, R. A., Lovegrove, S. and Davison, A. J.: DTAM: Dense tracking and mapping in real-time, *ICCV*, pp. 2320–2327 (2011).
- [14] Snavely, N., Seitz, S. and Szeliski, R.: Modeling the World from Internet Photo Collections, *IJCV*, Vol. 80, No. 2, pp. 189–210 (2008).
- [15] Taneja, A., Ballan, L., Pollefeys, M. and Pollefeys, M.: Image based detection of geometric changes in urban environments., *ICCV*, pp. 2336–2343 (2011).
- [16] CGAL: Computational Geometry Algorithms Library, <http://www.cgal.org>.
- [17] Vu, H.-H., Labatut, P., Pons, J.-P. and Keriven, R.: High Accuracy and Visibility-Consistent Dense Multi-view Stereo, *PAMI*, Vol. 34, No. 5, pp. 889–901 (2012).
- [18] Wendel, A., Maurer, M., Graber, G., Pock, T. and Bischof, H.: Dense reconstruction on-the-fly, *CVPR*, pp. 1450–1457 (2012).
- [19] Yu, S. and Lhuillier, M.: Incremental Reconstruction of Manifold Surface from Sparse Visual Mapping, *3DIMPVT*, pp. 293–300 (2012).