

# ハンドヘルドカメラを用いた歩行者の接地位置推定

○八木賢太郎 †, 長谷川邦洋 †, 杉浦裕太 †, 斎藤英雄 †

○ Kentaro YAGI † Kunihiro HASEGAWA † Yuta SUGIURA † and Hideo SAITO †

†: 慶應義塾大学, yagi@hvrl.ics.keio.ac.jp

近年, 様々なセンサを用いて人の動きを計測する手法が盛んに研究されている. その中でも「歩行」の計測は健康増進や病気の診断など様々な応用先があり重要視されている. しかしながら既存の計測手法の多くは特定のセンサが設置された空間内でしか使用できず, 長距離歩行の計測や生活する中で無意識に行っている自然な歩行を計測することは難しい. そこで提案手法では, より多様な環境で容易に使用可能な歩行計測手法として, 単一 RGB カメラによって撮影された歩行者の映像から 3 次元接地位置と歩行軌跡を推定する手法を提案する. 本論文では「屋外など, 広い空間での長距離歩行計測」そして「生活空間内での歩行計測」を想定し, 自由移動カメラを用いた計測手法と固定カメラを用いた計測手法の 2 つを提案する. 本手法のアルゴリズムは, 大きく背景 3 次元再構成と接地推定の 2 つの段階から成る. まず, Structure from Motion (SfM) を用いて背景の 3 次元モデルを再構築する. 画像間のマッチングを行う際に人検出領域内から検出された特徴点を除外することで, 背景物体のみを含む 3 次元モデルを再構成することができる. 続いて, SfM によって推定されたカメラ位置姿勢と, 畳み込みニューラルネットワーク (CNN) によって推定された人体関節座標を用いて, 計測対象人物が 3 次元空間中のどこに接地したかを推定する. 実験では, 移動カメラによって撮影されたシーン, 固定カメラによって撮影されたシーンそれぞれに対して本手法を適用し有効性を示す. <キーワード>歩行, 接地位置計測, 3 次元復元

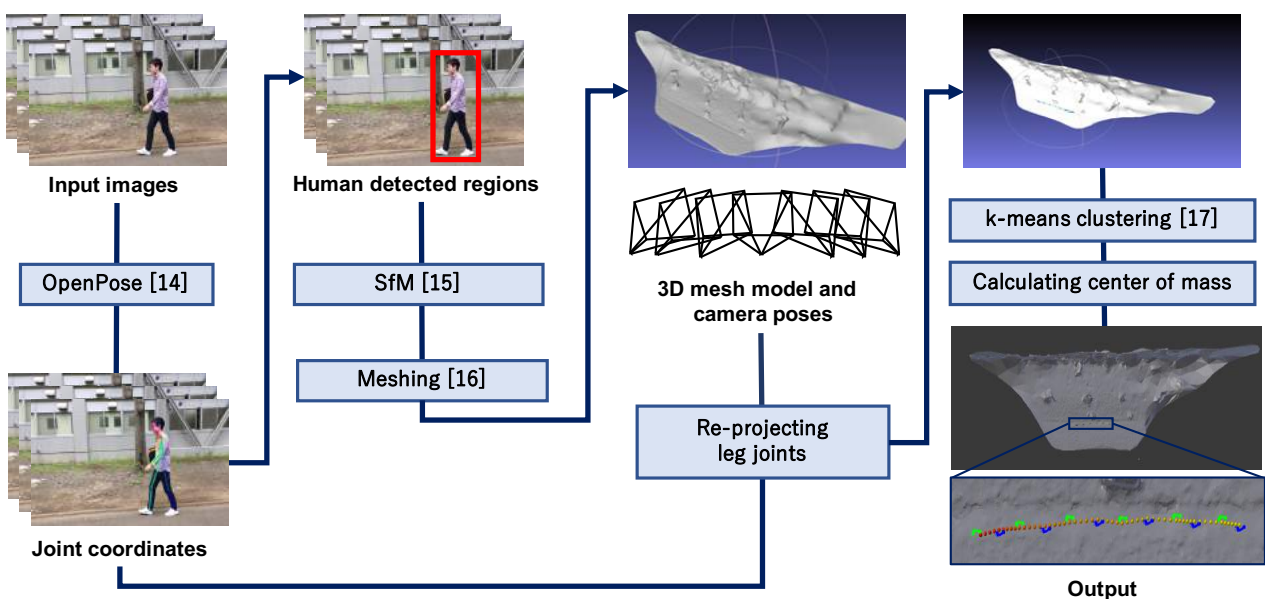


図 1 提案手法の流れ

## 1 はじめに

人の動きを計測することは行動予測、スポーツスキルの向上、疾患の発見など様々な分野に利用されている [1, 2]. 計測の対象となるシーンも多岐に渡るが、その中でも日常的な動作である「歩行」の計測は特に医療分野で重要視されており、人々の健康評価などにも利用されている [3, 4]. 歩行計測を行う方法として赤外線センサや圧力センサなどを用いた手法が挙げられるが、本論文では画像処理技術を用いた歩行計測に焦点を当てる。

画像処理を用いた研究の多くは「歩容認証」、即ち人の識別を目的としている。Jiwen らは歩行者のシルエット画像列を用いて映像中歩行者の性別を識別する手法を提案した [5]. Takamura らは、これまでの歩容認証手法が事前に想定された方向に歩く人物の画像を入力としていたのに対して、任意の方向に歩く人物の画像列から歩行者の認証を行う手法を提案した [6]. また服装の違いに対して頑健な手法として Hossain らの手法 [7], 歩行速度の違いに対して頑健な手法として Tsuji らの手法などが挙げられる [8]. こうした歩容認証手法は監視カメラで撮影された映像に写り込んだ人物を識別することを目的としており、カメラ位置が固定で、かつ解像度およびフレームレートが比較的低い映像を想定している。これに対し提案手法は歩行の計測を目的としている。3次元空間における歩行者の接地位置と軌跡を推定することを目的としており、3次元再構成を行うために十分な解像度を持つカメラを用いて歩行者の3次元接地位置と軌跡の推定を行う。

接地位置や歩行軌跡の計測を行う最も一般的な手法は赤外線センサを用いた手法である [9]. 赤外線反射材が塗布されたマーカーを多視点から撮影することで、マーカーの3次元位置を求めることができる。また床下に取り付けられた圧力センサを用いて歩行を計測する手法も存在する [10, 11]. Lee らは圧力センサを用いて、歩行者の歩幅、歩数、かかとがついてからつま先が離れるまでの時間を推定するシステムを提案した。Chen らは自宅など屋内での日常的な歩行を計測するために壁センサを用いて歩行速度と歩幅を測定するシステムを提案した [12]. Chen らの手法では微弱な信号を送信し、その信号の干渉度合いによって歩行位置や歩幅を推定する。GAITLAB は、赤外線カメラ、慣性センサ、圧力センサ、ワイヤレス筋電図センサなどの複数センサを組み

合わせ、より高精度かつ多岐に渡る情報の計測を可能にした [13]. ここで述べた手法のように画像処理技術を用いずとも多くの情報を計測できる手法は数多く存在する。しかしながら、こうした既存の計測手法が持つ問題点としてセンサが設置された限られた環境下でしか使用できないという問題点が挙げられる。そのため長距離歩行の計測や屋外での歩行を計測したい場合には向いていない。また日常生活空間と異なる環境での計測は被験者に計測されていることを意識させ、自然な歩行計測が難しい場合がある。そこで我々はハンドヘルドカメラを用いて歩行者の3次元接地位置と歩行軌跡を計測できる手法を提案する。本手法は既存手法に比べて広範囲での計測が可能である他、計測はカメラで歩行の様子を撮影することによって行われるため、被験者に計測を意識させず、より自然な歩行の計測が可能になると考えられる。

## 2 提案手法

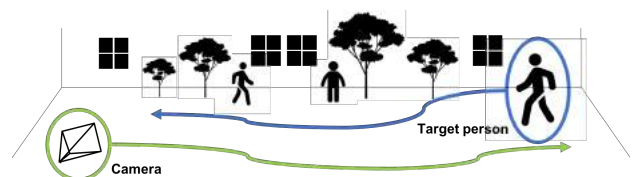


図2 対象とする撮影環境

本手法の目的は、RGBカメラによって撮影された映像から、歩行者の3次元接地位置と軌跡を推定することである。入力となる映像は図2に示すように歩行者と撮影者の両方が移動しながら撮影されることを前提としている。

まず、撮影された映像を入力として SfM により背景3次元モデルを復元する。事前に検出された人検出領域外の特徴点のみを用いてマッチングを行うことで背景領域のみを含む3次元モデルを復元することができる。続いて、SfMによって得られた3次元点群を各フレームに投影する。その後、各投影点とフレーム中歩行者の足関節位置の距離を求め、その値が閾値を下回った3次元点群へ投票を行う。全フレームでの投票終了後、投票数がある閾値を超えた3次元点を接地位置と判定する。これは、人が歩行を行う際には連続する数フレームで軸足が3次元空間上の同じ位置に存在するため、その間同じ3次元点が投票されると考えられるためである。提案手

法の流れを図 1 に示す。

## 2.1 対象人物のトラッキング

まず初めに計測対象人物の検出とトラッキングを行う。提案手法では画像中人物の関節検出モデルである OpenPose[14] を用いる。まず歩行者が映る最初のフレーム上の一点をクリックし、その点から各関節座標までの距離が最小の人物を計測対象とする。2 フレーム目以降では、前フレームでの対象人物の各関節座標と、現在のフレームで検出された各人物の各関節座標の距離を求め、その距離が最も小さい人物を計測対象とする。

## 2.2 背景 3 次元復元

続いて SfM を用いて背景 3 次元モデルの復元を行う。入力映像は歩行者だけでなく、撮影者も移動しながら撮影されたと仮定しているため、背景領域は複数視点から撮影される。SfM を行う際人検出領域から抽出された特徴点は除外する。これにより背景のみを含む 3 次元モデルを得ることができる。ここで人検出領域は OpenPose[14] によって推定された関節座標  $X = c1(x1, y1), c2(x2, y2), \dots, c18(x18, y18)$  の内、最大の  $x_i, y_i$  最小の  $x_i, y_i$  で定義される矩形とする。また SfM には Moulon らの方法を用いる [15]。SfM によって疎な点群を取得した後、Jancosek らの手法を用いて点群のメッシュ化を行う [16]。

## 2.3 接地位置推定

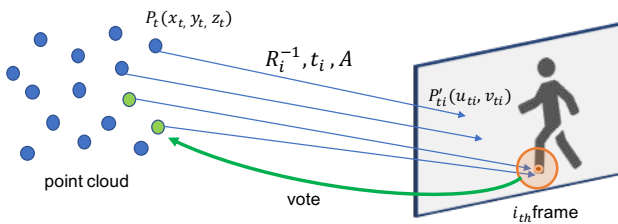


図 3 移動カメラを用いた手法における 3 次元点群への投票方法

背景 3 次元モデルの復元後、歩行者の接地位置推定を行う。この時カメラキャリブレーションによって事前に求めておいたカメラの内部パラメータと SfM によって求めたカメラの外部パラメータを用いる。ここで、カメラの内部パラメータを  $A$  とし、フレーム  $i$  におけるカメラの回転行列と並進ベクトルをそれぞれ  $R_i, t_i$  とする。またフレーム  $i$  に映る、計測対象人物の右足関節を  $r_i$ 、左足関節を  $l_i$  とする。まず、全ての 3 次元点群を各フ

レームの画像平面に投影する。ここで、 $t$  番目の 3 次元点群を  $P_t$ 、フレーム  $i$  における  $P_t$  の投影先の点を  $P'_{ti}$  とする。この時  $P'_{ti}$  は以下の式で表される。

$$P'_{ti} = A (P_t - t_i) R_i^{-1} \quad (1)$$

続いて、画像平面における投影点  $P'_{ti}$  と  $r_i, l_i$  の距離  $D_{ri}, D_{li}$  を計算する。ここで  $D_{ri}, D_{li}$  がしきい値  $th_d$  未満である時、 $P_t$  に投票する。 $th_d$  は経験的に 10 とする。ここで投票された点はそのフレームにおける接地位置候補と見なされる。投票処理を図 3 に示す。

次に、どの候補点が接地位置であるかを推定する。各候補点を接地位置と判定するかは投票数によって決定する。人が歩行する時、接地している足の 3 次元位置は連続する数フレームで同じである。その為、接地している足の近傍領域に対応する 3 次元点群は連続した数フレームで投票される。従って、他の点と比べ投票数が多い点は接地点であると考えられる。ここで  $th_v$  は経験的に 8 とした。ここで用いた  $th_v$  の適切な値はカメラのフレームレートと対象者のピッチによって異なる。その為、本手法のユーザは撮影シーンに合わせた適切な  $th_v$  を決定する必要がある。

ここまでで 3 次元点群中のどの点が接地点かは分かっているが、どの点が  $N$  歩目の接地位置に対応するかは未知である。そこで接地点と判定された点群に対して k-means 法を用いてクラスタリングを行う [17]。この時クラス数  $k$  は歩数に対応し、その値は入力映像の歩数を数えることで得る。クラスタリング後、各クラスの代表点を決定する。代表点は同じクラスに分類された点の重心とし、この代表点を用いて歩幅を求めることができる。また映像中に存在するサイズ既知のオブジェクトを用いて 3 次元モデルのスケール合わせを行うことで絶対スケールの歩幅を求めることもできる。最後に歩行軌跡の求め方について説明する。軌跡を求めるには各クラスの接地時刻が分かれば良い。接地時刻は同じクラスに分類された各点における各投票点の投票時刻の平均とする。

## 2.4 固定カメラを用いた接地位置・軌跡推定

ここまでは自由移動カメラによって撮影された映像を対象としていたのに対して、本節では固定カメラで撮影された映像から 3 次元接地位置と歩行軌跡を推定す



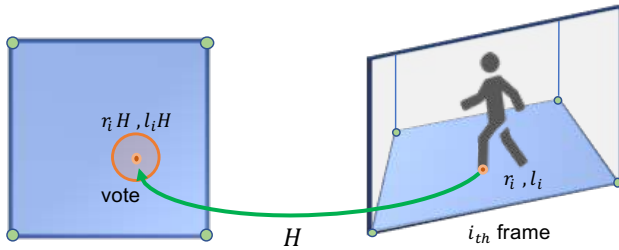


図 4 固定カメラを用いた手法における床平面画像への投票方法

る手法について説明する. 本手法は日常生活における歩行のモニタリングに用いられることを想定しており, 部屋にカメラを設置するだけで居住者の歩行計測を行うことができる.

自由移動カメラを用いた手法との最も大きな違いは, 視点が固定されているために SfM によって背景 3 次元モデルを復元できない点である. そこで本手法では平面射影行列によって画像平面と 3 次元空間内の床平面の対応関係を得る. そのため本手法では計測環境の床が完全な平面であると仮定している. また本手法では事前にカメラの内部パラメータを求めておく必要はない.

画像平面と床平面の対応を示す平面射影行列を  $H$  とする.  $H$  は画像平面と床平面の上の 4 組の対応点を手動で与えることで求める. 接地点を決定する際は自由移動カメラによって撮影されたシーンと同じように OpenPose[14] によって得られる足関節位置を用いる. まず各フレームにおける足関節位置を  $H$  によって床平面に投影する. 続いて投影点を中心とする半径  $r$  以内に存在するピクセルに投票する. 最後に床平面画像中の各ピクセルの投票数を数え, その値が閾値  $th_v$  を超えたピクセルを接地点とする. 本手法における床平面画像への投票方法を図 4 に示す.



図 5 入力画像の例

### 3 結果

本節では背景が異なる 3 つのシーンに対して提案手法を適用し有効性を示す. Scene 1 と Scene 2 は自由移動カメラによって撮影された映像を入力としており Scene 3 は固定カメラで撮影されたシーンを入力としている. Scene 1 は校庭で撮影され, 撮影者から歩行者までの距離は約 10 m であった. またフレーム数は 486 である. Scene 2 は背景に建物が存在する道路上で撮影され, 撮影者から歩行者までの距離は約 5 m であった. またフレーム数は 560 である. Scene 3 は室内で撮影され, カメラは部屋の隅に固定されている. フレーム数は 200 である. 撮影には解像度  $1920 \times 1080$ , シャッター速度 30 fps の iPhone 6S を用いた. 各シーンにおける入力画像例を図 5 に示す.

Scene 1 と Scene 2 では接地点・軌跡に加えて計測した 12 歩の歩幅を絶対スケールで推定した. Scene 1, Scene 2 共に歩行者は床上のメジャーに合わせて均等な歩幅で歩いた. スケール合わせを行う際には背景に存在するサイズ既知の物体を用いて, 最初のストライドが正しく推定されたと仮定しスケール合わせを行なった. Scene 1 では歩行者が 80 cm の歩幅で歩いたのに対して, 推定された歩幅の平均は 81 cm であった. Scene 2 では, 歩行者が 60 cm の歩幅で歩いたのに対して, 推定された歩幅の平均は 59 cm であった. 従って推定された歩幅と実際の歩幅の差の平均は約 1 cm であった.

Scene 1 と Scene 2 の結果は図 6 に, Scene 3 の結果を図 7 に示す. 各画像において青色で示された領域が左足の接地点, 緑色で示された領域が右足の接地点を示す. 黄色と赤色の点は歩行の軌跡を示し, 時間が経過するにつれて点の色が黄色から赤に変化している. 図 6, 図 7 を見るとほぼ均等な間隔で交互に接地していることが分かる. このことから相対的な接地点は大よそ正しいと考えられる. しかしながら絶対スケールの歩幅は必ずしも正しいとは言い切れない. Scene 1 と Scene 2 の推定平均歩幅と実際の歩幅の差は約 1 cm であったが,  $th_d$  や  $th_v$ , スケール合わせの方法, 歩行者が歩く際のメジャーとのズレなど結果に影響を与える要因は様々な存在している. 従って, 実験結果が信頼できるものであるかどうかを判断することはできず, より多様なシーンにおいて提案手法を評価する必要がある.

また推定誤差の原因として OpenPose[14] によって推

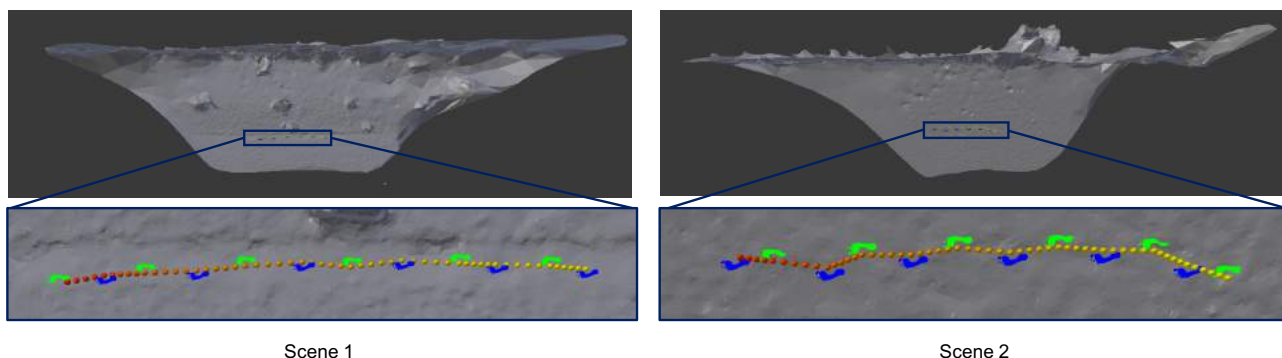


図 6 Scene 1, Scene 2 における 3 次元モデルと接地点・軌跡の可視化結果

定された関節位置の誤差があげられる。OpenPose[14]の結果をフレームごとに見ると、推定された座標が実際の足関節と大きくずれるシーンがいくつか含まれていた。また、接地位置検出のために用いている足関節は実際に地面に接地している領域よりも足首に近く、推定接地位置はカメラ、脚の関節、および床平面のなす角度に応じてズレると考えられる。正確な接地位置を推定するには、OpenPose[14]によって与えられる足関節位置ではなく、地面と接地している箇所の画像座標を用いる必要がある。

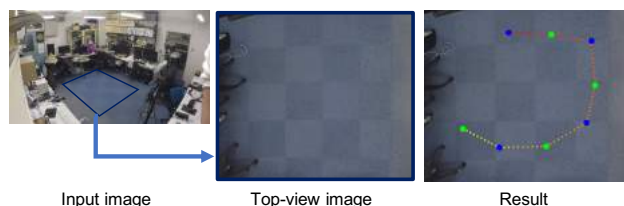


図 7 Scene 3 における接地点・軌跡の可視化結果

## 4 結論

単一 RGB カメラを用いて歩行者の 3 次元接地位置と軌跡を推定する手法を提案した。既存の計測手法の多くは、センサが設置された限られた環境でしか計測を行うことが出来ないため、長距離の歩行計測ができない、自然な歩行が計測できないといった問題があった。それに対し我々は歩行者の映像から 3 次元設置位置、軌跡を求める手法を提案することで、計測可能環境を拡張すると共に、自然な歩行の計測を可能にした。実験では異なる 3 シーンに対して提案手法を適用し本手法の有効性を確認した。

## 謝辞

本研究は、AIP-PRISM, Japan Science and Technology Agency (JPMJCR18Y2) の支援を受けたものである。

## 参考文献

- [1] Shakya Subarna, Sharma Suman, and Basnet Abinash. Human behavior prediction using facial expression analysis. In *International Conference on Computing Communication and Automation (ICCCA)*, pp. 399–404, 2016.
- [2] Hassan Ghasemzadeh, Vitali Loseu, Eric Guenterberg, and Roozbeh Jafari. Sport training using body sensor networks: A statistical approach to measure wrist rotation for golf swing. In *Body Area Networks*, p. 2, 2009.
- [3] Stacy Fritz and Michelle Lusardi. White paper: walking speed: the sixth vital sign. *Journal of Geriatric Physical Therapy (JGPT)*, Vol. 32, No. 2, pp. 2–5, 2009.
- [4] Stephanie Studenski, Subashan Perera, Kushang Patel, Caterina Rosano, Kimberly Faulkner, Marco Inzitari, Jennifer Brach, Julie Chandler, Peggy Cawthon, Elizabeth Barrett Connor, et al. Gait speed and survival in older adults. *Journal of the American Medical Association (JAMA)*, Vol. 305, No. 1, pp. 50–58, 2011.
- [5] Lu Jiwen, Wang Gang, and Moulin Pierre. Human identity and gender recognition from gait sequences with arbitrary walking directions. *IEEE*

- Transactions on Information Forensics and Security (TIFS)*, Vol. 9, No. 1, pp. 51–61, 2014.
- [6] Takemura Noriko, Makihara Yasushi, Muramatsu Daigo, Echigo Tomio, and Yagi Yasushi. On input/output architectures for convolutional neural network-based cross-view gait recognition. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, pp. 1–1, 2018.
- [7] Md Altab Hossain, Yasushi Makihara, Junqiu Wang, and Yasushi Yagi. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, Vol. 43, No. 6, pp. 2281–2291, 2010.
- [8] Akira Tsuji, Yasushi Makihara, and Yasushi Yagi. Silhouette transformation based on walking speed for gait identification. In *Computer Vision and Pattern Recognition (CVPR)*, pp. 717–722. IEEE, 2010.
- [9] Ltd. OptiTrack Japan. Optitrack. <https://www.optitrack.co.jp/>, 2018. Accessed: 2018-12-10.
- [10] Lee Middleton, Alex A Buss, Alex Bazin, and Mark S Nixon. A floor sensor system for gait recognition. In *Automatic Identification Advanced Technologies (AIAT)*, pp. 171–176. IEEE, 2005.
- [11] Philipp Leusmann, Christian Mollering, Lars Klack, Kai Kasugai, Martina Ziefle, and Bernhard Rumpel. Your floor knows where you are: sensing and acquisition of movement data. In *Mobile Data Management (MDM)*, Vol. 2, pp. 61–66. IEEE, 2011.
- [12] Chen Yu Hsu, Yuchen Liu, Zachary Kabelac, Rumien Hristov, Dina Katabi, and Christine Liu. Extracting gait velocity and stride length from surrounding radio signals. In *Conference on Human Factors in Computing Systems (CHI)*, pp. 2116–2126, New York, NY, USA, 2017. ACM.
- [13] BTS Bioengineering. Gaitlab. <http://www.btsbioengineering.com/products/bts-gaitlab/>, 2018. Accessed: 2018-12-10.
- [14] Cao Zhe, Simon Tomas, Wei ShihEn, and Sheikh Yaser. Realtime multi-person 2d pose estimation using part affinity fields. In *Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [15] Pierre Moulon, Pascal Monasse, and Renaud Marlet. Adaptive Structure from Motion with a contrario model estimation. In *Asian Conference on Computer Vision (ACCV)*, Vol. 7727 of *Lecture Notes in Computer Science*, pp. 257–270, Daejeon, South Korea, November 2012. Springer.
- [16] Jancosek Michal and Pajdla Tomas. Exploiting visibility information in surface reconstruction to preserve weakly supported surfaces. *International Scholarly Research Notices (ISRN)*, Vol. 20, , 2014.
- [17] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, Vol. 28, No. 2, pp. 129–137, 1982.