

都市など広域空間の効率的モデリングおよびレンダリング手法について

川崎 洋^{†1} 永塚 遼^{†2} 小野 晋太郎^{†2}
栗林 宏輔^{†3} 子安 大士^{†3}
前川 仁^{†3} 池内 克史^{†2}

著者らの研究グループはこれまで、建物など市街の建造物の形状やテクスチャなどの情報を取得し、これをコンピュータグラフィクス(CG)により現実感豊かに計算機内に再現する研究や開発を行ってきた。本発表では、これまで研究グループが取り組んできた手法の概略を述べるとともに、最近の研究成果である、広域空間の情報の効率的な取得方法や、テクスチャ品質の向上手法、ドライビングシミュレータとしての応用方法などについて紹介する。

Introduction of City Modeling Project —capturing, modeling and rendering of large scale scene—

HIROSHI KAWASAKI,^{†1} SHINTARO ONO,^{†2}
RYO NAGATSUKA,^{†2} KOUSUKE KURIBAYASHI,^{†3}
HIROSHI KOYASU,^{†3} HITOSHI MAEKAWA^{†3}
and KATSUSHI IKEUCHI^{†2}

More than 10 years, authors research group has tried to achieve an efficient acquisition of shape and texture of buildings in the urban city, modeling of them and photo-realistic rendering techniques by computer graphics. In this presentation, brief explanations of recent contributions of the authors work are presented, such as acquisition of large 3D city by integrating multiple car-mounted videos, efficient technique to remove obstacles from building textures, and photo-realistic driving simulator using omni-images.

1. はじめに

近年、実世界データ、特に地図やそれに付随する情報を利用するサービスやアプリケーションが、インターネット上を中心に幅広く利用されるようになってきている。我々の研究グループはこれまで、このような実世界・地形情報の中でも特に市街を構成する主な建造物である建物の情報(主に形状やテクスチャ)に注目し、これらを効率的に取得し、写実的に表示するための研究を行ってきた¹⁾。このように広域な空間の情報を効率よく取得し、モデリングし表現するためには多くの課題や問題があるが、特に以下の点が重要である。

- (1) 広い空間の情報を効率よく取得するための手法や機器、およびその自動化アルゴリズム
- (2) 複雑な形をした建物の3次元情報の取得精度の向上
- (3) 天候や日照の変化、歩行者や街路樹などの前景物の影響により安定した取得の難しい、建物テクスチャの品質向上
- (4) 時間変化(車両追跡のような短い時間経過から、建物の経年変化など長い時間経過まで)に伴い変化する情報の取得

これらの問題点に対して我々は以下のような解決方法を提案し取り組んできた。

- (1) 車載カメラの流し撮りによる効率的なデータ収集に基づく情報取得手法
- (2) インターネット上にアップされた多数のデータを効率良く利用するための方法
- (3) 高品質なテクスチャ情報の取得のため、窓ガラスに写り込んだ反射像を除去する手法
- (4) 歩行者や街路樹、走行車両などを除去し、建物テクスチャのみを取得する手法
- (5) Structure from motion に代表される、レンジセンサを用いない形状取得手法
- (6) イメージベースレンダリングによる形状によらない写実的表現手法

以下ではこれまで提案してきた解決手法のうち、最近の成果を中心に紹介する。

2. 複数の車載カメラからの運動および形状の同時復元手法

インターネット上などに公開されている、車載カメラによる映像を統合し、広域空間を構

^{†1} 鹿児島大学

Kagoshima University

^{†2} 東京大学

Institute of Industrial Research, University of Tokyo

^{†3} 埼玉大学

Saitama University

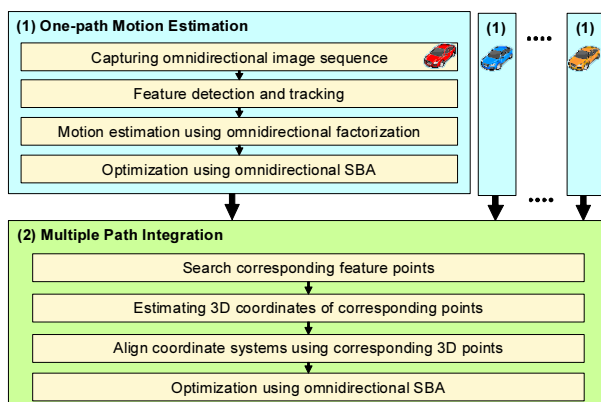


図 1 Outline of the proposed method

築することが出来れば効率良く広域の市街モデルを構築することができる。ここでは、これを実現する手法を紹介する。

本手法の概要を Fig. 1 に示す。手法は大別して 2 段階で構成されている。Step (1) は単独の車載全方位カメラによって撮影されたビデオに対する処理であり、別車両の車載カメラ動画、あるいは同車両で別時刻に撮影した動画像に対して、それぞれ独立に行われる。

まず Step (1) によって、車両運動が全方位ビデオを元に推定される。ここでは SfM の手法の 1 つであり、一般的な透視投影画像列を扱う因子分解法を用いる。その際、全方位画像に対して頑健な処理が実現可能な拡張された因子分解法を提案した。

次に Step (2) において、個別に復元した複数の 3 次元構造とカメラパスを 1 つの座標系へ統合し、改めてバンドル調整を適用する。その結果、密度や正確さが向上し、スケラビリティをもつ広域な仮想都市モデルを構築する事が出来る。以下に手法の流れを順に説明する。

2.1 特徴点抽出および追跡

最初に撮影した画像列の特徴点抽出および追跡を行う。特徴点抽出には SIFT²⁾ を用い、特徴量を比較することで初期フレームから最終フレームまで対応が続いている点を確保する。しかし誤対応が生じる場合もあるため、エピポラ拘束と RANSAC³⁾ を利用し、外れ点の除外処理を行う。これは対応付けされた特徴点から 5 組をランダムに選択し、5-point algorithm⁴⁾ によって前後フレームによるエピポラ平面を推定し、この平面に近い点が一

定数以上の時、推定されたエピポラ平面は正しいとし、それに当てはまらない点を除外することで実現される。

2.2 因子分解法によるカメラ運動推定

因子分解法⁵⁾ は、多視点の画像から 3 次元形状とカメラ位置姿勢を推定する手法である。特に今回は透視投影カメラを想定した因子分解法^{6),7)} を用いた。全方位画像の場合は画像を分割し、それぞれに因子分解法を適用する。これを複数繰り返す、統合することで全方位の形状と位置姿勢の初期値を推定する。

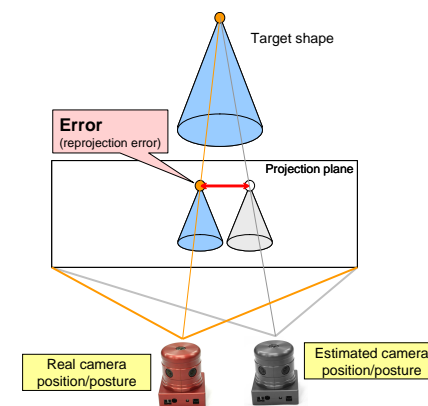


図 2 Bundle adjustment minimizing the reprojection error

2.3 複数系列データの統合のための共通特徴点の検出

ここまでの処理で、1 系列の全方位画像列から周辺の 3 次元形状とカメラパスの初期値が同時に推定される。そこで、こうして得られた複数の推定データを統合する。このため、最初に複数映像に共通の特徴点を抽出する。

通常のカメラでは視野が狭いため、視線の方向や画角、焦点距離などの違いにより、対応点を安定して求めることは容易ではない。本論文では、一つの全方位カメラ映像をリファレンスシーケンスとして用いることで、常に対応点が得られやすい利用方法を想定する。これは一台の全方位カメラが主要道路を走行することで実現でき無理の無い前提であり、常に対応点が得られることは、複数経路の統合では特に重要である。

特徴量としては、全方位画像を円筒座標系に変換し、円筒座標系で SIFT²⁾ を計算して用

いた．実際の処理としては，各系列の全方位画像列から SIFT 特徴量を求めておき，その中で多く検出された特徴量を共通特徴点候補としてデータベースに保存する．系列同士のマッチングは共通特徴点データベース内で類似の特徴量を探索することで実現する．

2.4 複数系列の同時バンドル調整

こうして得られた複数経路の統合は

- (1) 3次元特徴点群同士の幾何的なマッチングによる粗い位置合わせ
 - (2) バンドル調整による高精度位置合わせ
- という2ステップで実現される．

2つの画像系列間の対応点の組 X_i, X'_i が得られると，それらの相対座標は次の方程式の最小化によって知ることができる．

$$\min_{s,R,t} \sum_i |X_i - s(RX'_i + t)|^2 \quad (1)$$

s, R, t はそれぞれ相対スケール，姿勢，位置を示す．これを用いることで全ての系列座標系を1つのワールド座標系へ統合する．さらに，この1つのワールド座標系で，改めて全ての特徴点に対し再投影誤差を最小化するバンドル調整を行う．これにより統合後の1座標系に解が得られる．

2.5 複数経路データによる広範囲な3次元復元実験結果

長距離を撮影したビデオ映像を，短い系列に分割して3次元復元後，それらを統合する実験を行った．ここでは実際に全方位カメラを車載し，都市を約150m撮影走行して得られたビデオを20の短いビデオに分割した．Fig. 3は推定した短い系列を統合した結果である．

各系列の推定結果はスケールや座標系が異なるが，統合後全体にバンドル調整を行うことでそれぞれに矛盾の無い解が得られた．これにより，連続したビデオ映像については長距離を撮影したものであっても，本手法で復元可能であることが確認できた．また20の系列に分割したにも関わらず，総合結果はほぼ直線状になっており，ロバストな処理が実現できている事が分かる．

更に多系列のデータに対応付けし，密な3次元を復元する実験を行った．この実験では撮影車両で都市部の公道を走行し，同じ区間を数回程度走行し，得られた複数の画像系列にSfMを適用すると同時に，各ビデオからSIFTを利用して共通特徴点を抽出している．3系列(計1113フレーム，実距離約220m)のデータを統合して得た密な3次元の復元結果をFig. 4に示す．カメラ以外に特別な装置を使用せず，平面性などの事前知識を利用することなく，より広範囲なシーン復元に成功している．

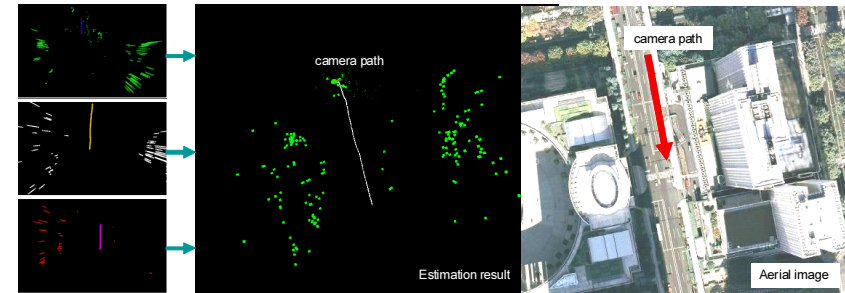


図3 Result of long sequence integration

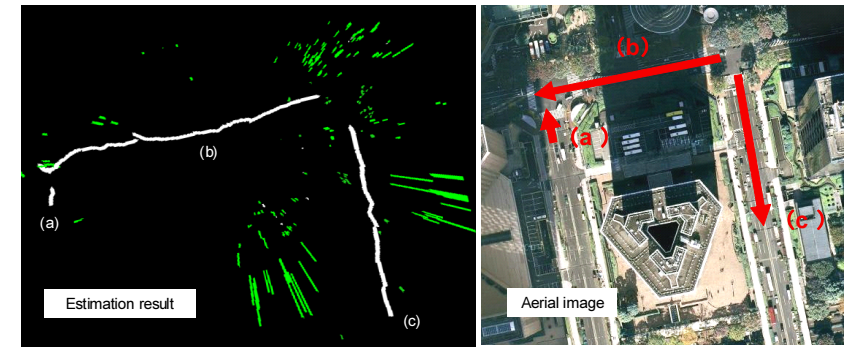


図4 Result of multiple sequences integration

3. テクスチャクオリティの向上

自動車で移動しながらテクスチャの取得を行った場合，本来の目的である建物テクスチャ以外のものが多く映り込むことになる．例えば，図5(a)に示すように，Google Earthなどの3Dモデル表示ソフトでは，ビルの壁面に貼り付けられているテクスチャにビルの壁面以外のものが映り込んでおり，違和感を感じさせる．また，図5(b)に示すように，Google Street Viewでは取得した画像に歩行者や車のナンバーが映り込むプライバシーの問題がある．

このような映り込みを除去する手法の一つとして，Floresらは画像上で歩行者を検出し，他の画像からその領域をコピーする手法を提案した⁹⁾．歩行者全体を除去しているものの，

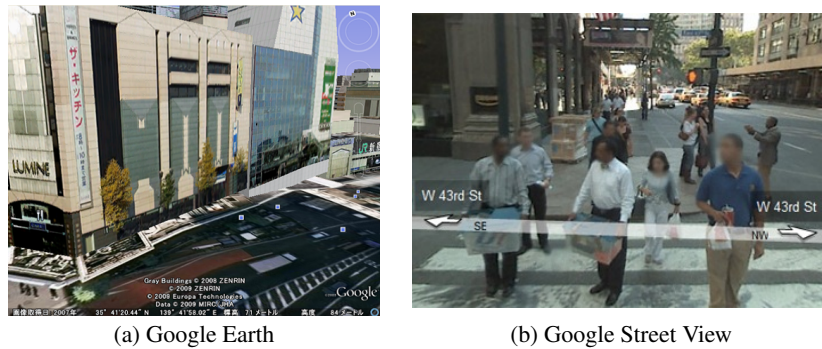


図5 都市モデリングにおけるテクスチャの問題。
Fig. 5 Problem of the texture of the city modeling.

背景の幾何情報を考慮していないため、除去結果は違和感を感じさせる。

そこで、我々は歩行者や街路樹などの映りこみを前景障害物と定義し、これを除去することで取得の効率性を損なわずに高質なテクスチャを取得する手法を提案した¹⁰⁾。撮影対象であるビル壁面や道路などの都市建造物が平面として近似できることを利用し、広域な都市空間を撮影した画像列から前景障害物を指定することなく除去し、高質な都市建造物のテクスチャを取得する。

高質なテクスチャ取得手法は、まず、車載カメラによって都市建造物を中心に移動しながら撮影し、入力画像列を取得する。次に、入力画像列間において特徴点を追跡し、特徴点群の軌跡を解析する事で都市建造物平面を推定する。求めた平面毎に時空間画像解析によって前景障害物除去する。最後に、各平面における除去結果を統合しシーン全体の除去結果を得、都市建造物のテクスチャとする。

3.1 高質なテクスチャ取得手法

特徴点群の軌跡から都市建造物平面を推定する。まず、撮影した各入力画像に AffineSIFT¹²⁾ によって特徴点抽出と対応点探索を全画像間で行い、特徴点群の軌跡を得る。次に、求めた特徴点群の軌跡を解析することによって都市建造物平面を推定する。ここで推定する平面とは同一の平面に属する特徴点の集合である。これは、画像間における同一平面同士¹³⁾の射影変換による再投影誤差を利用する¹⁴⁾。

各平面毎に時空間画像解析を利用し、前景障害物を除去する。まず、時空間画像解析を容易にするため、3.1 節で求めた平面を基に入力画像列を射影変換し、位置合わせ画像列を作成する。これにより、位置合わせ画像列上では、平面の見えが変らないのに対して前景障

害物が移動する。次に、座標毎に位置合わせ画像列を走査し、色を取得する。取得した色の RGB を 3 次元のベクトルとして Vector Median Filter (VMF)¹⁵⁾ に入力し、最も中間的なベクトルを得る。これによって得たベクトルを都市建造物の色として採用し、対象とした平面における前景障害物除去結果とする。

各平面における前景物除去結果を統合し、シーン全体の除去結果を生成する。しかし、ここまでで得ている平面とは同一平面に属する特徴点群のみであり、平面領域については把握していない。そこで、シーン全体の除去結果は座標毎に各平面の除去結果から色を選択する。選択方法は各平面における除去結果の RGB 値とその座標における位置合わせ画像列の全フレームの RGB 値を比較し、色空間のユークリッド距離が閾値内の数をカウントし、最も多い平面 RGB 値を選択する。

3.2 車載カメラ映像による実験結果

車載カメラによる実験結果を図 6 に示す。背景のビル群を平面と仮定し、手前に存在する街路樹や高架橋を前景障害物として、前景障害物除去実験を行った。図 6 の左列から順に入力画像、入力画像に背景平面推定点をプロットした画像、基準画像 (11 フレーム目) の背景平面に対して入力画像を射影変換した位置合わせ画像、VMF によって前景障害物除去した結果画像である。平面点プロット画像では、平面推定の処理によって背景の平面領域上に存在する特徴点が正しく選択されているのがわかる。結果画像では、前景障害物である街路樹や高架橋の大部分が除去され、高品質の建物テクスチャが取得できる。

4. 市街の実データを用いたアプリケーション

我々のグループで開発してきた実写画像ベースの IBR とポリゴンベースの MBR を組み合わせ、ドライビングシミュレータについて紹介する。

本手法では、写実的な再現と、固有空間法による効率的な圧縮の両方を同時に実現することが可能である。我々が開発したドライビングシミュレータシステムの全体像を図 7 に示す。遠距離にある建物、空の部分は実世界のモデルコースを撮影して得た画像データを処理して描かれ、近距離にある道路、車、標識などの部分は既存製品の一部をそのまま活用して、それぞれ独立に描かれる。また、その間の中距離にある看板、電柱、高架橋といったものに関しては画像処理による除去を行う。

画像ベースの Image-Based Rendering (IBR) とポリゴンベースの Model-Based Rendering (MBR) の見えは、コンポジットと呼ばれるハードウェアにより、それぞれの奥行き値を判定することで合成され、ユーザはそれを見て運転操作を行う。運転操作はホストに帰還

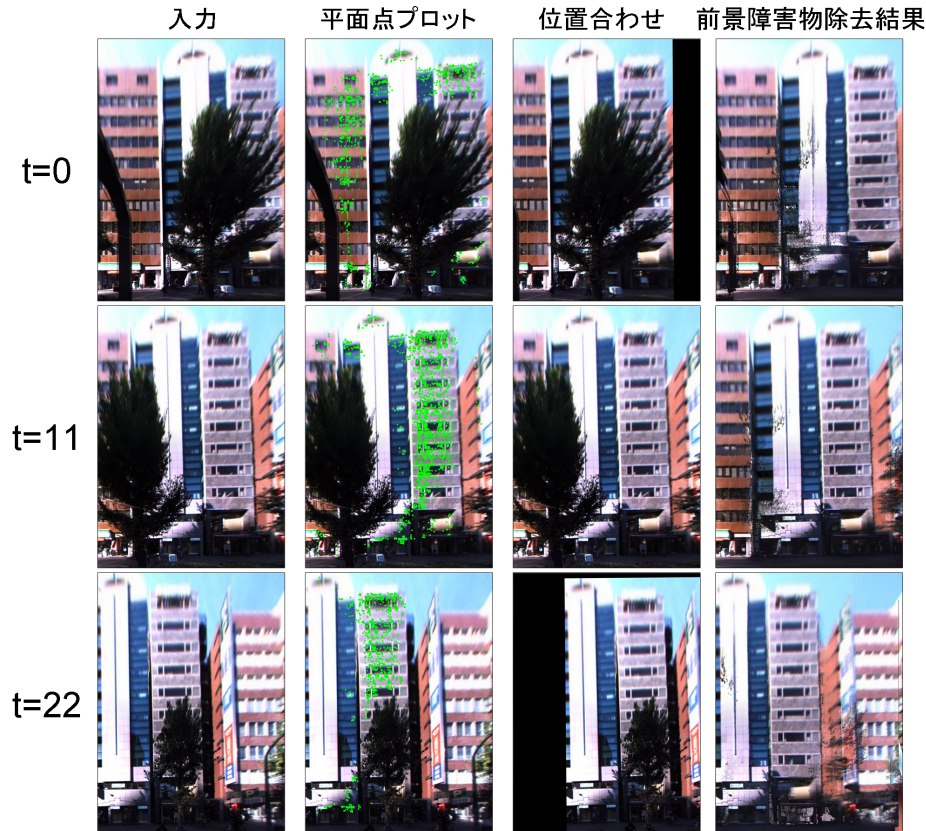


図 6 車載カメラ映像実験：入力画像，平面点プロット画像，位置合わせ画像，前景障害物除去結果。
Fig. 6 On-vehicle video experimental: input images, plane point images, registration images, results of foreground obstacle removal.

され，ホストから受け取った自己車両の位置・姿勢は次の描画処理に反映される。

また，IBR のデプスマップはレンダリング結果とは別に独立して，道路構造に合わせた左右の板と天井板の 3 枚の板で構成されるシーンを描画し，その奥行き値を利用する。画像のレンダリング結果と同様に，正面部分に奥行き値の板を置くと，正面方向から来るポリゴンの車が，画像ベースの部分の何も無かった場所から飛び出してくるため，IBR のデプスマップでは正面部分に板を配置しないようにした。

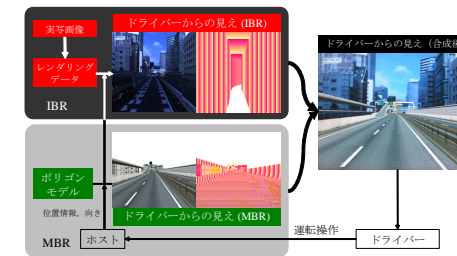


図 7 提案するドライビングシミュレータシステムの全体像

本実験で用いるドライビングシミュレータは図 8 のような，ミニバン実車両のカットボディを使用する。実車両と同等の電動パワーステアリング装置やブレーキ装置を搭載しており，臨場感の高い運転をすることができる。また，速度を上げるとハンドルが硬くなるといった運転挙動も再現されている。

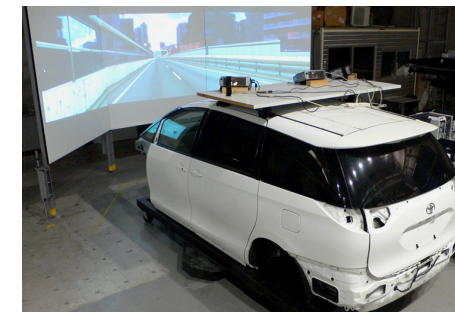


図 8 カットボディのシミュレータ

IBR と MBR のレンダリングはそれぞれ別の PC で行う。そのためにはまず，運転席からハンドルのステアリング角度とアクセル，ブレーキの踏み込み量がホスト PC の方に送られる。ホスト PC は現在位置と方向，速度と送られてきたステアリング角度，アクセル，ブレーキの踏み込み量を元に次の車両位置と方向，速度を計算し，決定する。決定された車両位置と方向をバケットにして各レンダリング PC に送り，IBR と MBR の描画の切り替えを行う。

その際，IBR を出力する PC と MBR を出力する PC は RGB チャンネルと Z チャンネルをコ

ンボジッタに入力する．コンボジッタ内で IBR の Z チャンネルと MBR の Z チャンネルを各ピクセルごとに比較し，デプスが手前の方の RGB チャンネルを出力する．これによって IBR と MBR を合成し，各プロジェクタよりその合成映像を出力する．

以上の合成工程を図にまとめると図 9 のようになる．

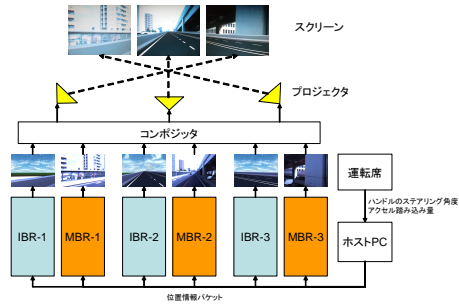


図 9 ドライビングシミュレータシステム設計図

以後，本ドライビングシミュレータに用いられている自由視点レンダリングについて説明する．

自由視点からの見えは，走行経路上で撮影した全方位画像の一部をつなぎ合わせることで合成する．例えば図 10 では，印の地点は撮影時には走行していないが，そこからの左方への見えは，それぞれ時刻 t_1, t_2, t_3 における全方位画像の左前方，左方，左後方の光線を参照してつなぎ合わせるにより合成することができる．これにより，車線変更効果を再現することができる．

広域空間を再現するために，図 11 のように正面と側面の板を仮定し，それぞれにテクスチャマッピングすることで広域空間のレンダリングを行う．側面においては表面光線空間法に基づいた自由視点レンダリングを行う¹⁶⁾．表面光線空間法では，物体の表面から光線が出ているものとし，物体表面上で光線空間を定義している．

また，進行方向である正面に関しては，物体の表面を定義できないため，全方位画像の進行方向部分を利用することで対応する．

図 12 は自由視点レンダリングをする様子を真上から見た図である．

自由視点レンダリングにおいては，ドライバーのヘディング角に応じて水平視角を変化させるだけでなく，ドライバー（視点）が撮影経路から左右にずれたときに，その地点におけ

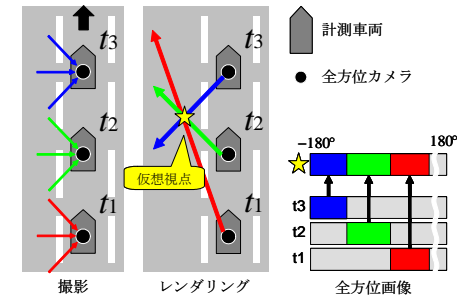


図 10 新しい視点からの見え生成

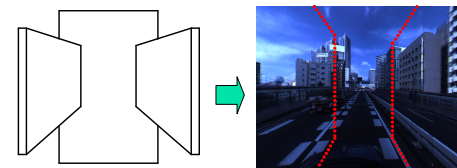


図 11 テクスチャマッピング

る見えを生成する．

5. ま と め

我々がこれまで取り組んできた，効率的な街のモデリング手法について，最近の成果である，1．複数映像の統合による広域空間の自動生成手法，2．前景物の除去による高品位建物テクスチャの取得手法，3．写實的ドライビングシミュレータの構築，の3つについて紹介した．今後は，Web での利用を想定したより効率的なデータ形式や圧縮手法，広域空間の高精度な 3 次元情報の自動取得手法などが重要と考えている．

参 考 文 献

- 1) 子安大士, 古屋大和, 深澤龍一郎, 川崎洋, 前川仁, 小野晋太郎, 池内克史, “サブピクセル推定を用いた全方位ステレオ視による高精度な 6 自由度 slam,” 第 2009 巻, pp. 19–24, 2009.
- 2) D.G. Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, vol.60, no.2, pp. 91–110, 2004.

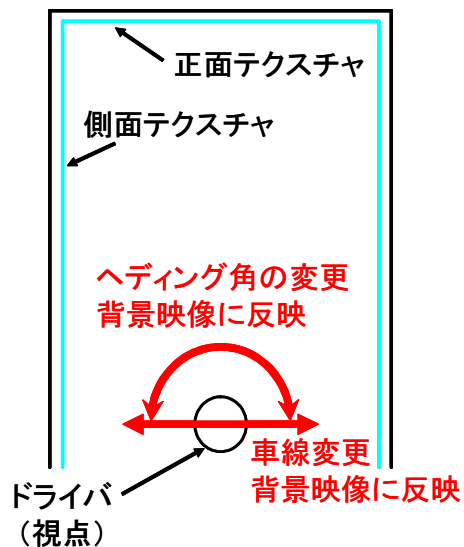


図 12 自由視点レンダリング

動カメラ映像からの前景物とテキスト面の自動分離”, 第 13 回画像の認識・理解シンポジウム (MIRU2010), IS1-32, pp.255-262, July 2010.

- 11) B. Lucas and T. Kanade. “An iterative image registration technique with an application to stereo vision,” Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI’81), pp.674-679, 1981.
- 12) J. M. Morel and G. Yu, “ASIFT: A new framework for fully affine invariant image comparison”, Society for Industrial and Applied Mathematics Journal on Imaging Sciences, Vol. 2, No. 2, pp. 438-469, 2009.
- 13) 山口拓真, 川崎洋, 古川亮, 中山利宏, “画素単位のレジストレーションによる 3 次元形状物体画像の超解像処理”, 第 12 回 画像の認識・理解シンポジウム (MIRU2009), IS2-54, pp.1223-1230, July 2009.
- 14) 矢口陽一, 田中正行, 奥富正敏, “超解像処理のための複数モーションに対応したロバストかつ高精度な位置合わせ手法”, 電子情報通信学会論文誌, Vol.J92-D, No.11, pp.2033-2043, 2009.
- 15) J. Astola, P. Haavisto, and Y. Neuvo, “Vector Median Filters”, Proc. of the IEEE, Vol.78, No.4, pp.678-689, Apr 1990.
- 16) D. Wood, D. Azuma, W. Aldinger, B. Curless, T. Duchamp, D. Salesin and W. Stuetzle: “ Surface light fields for 3d photography, ” ACM SIGGRAPH 2000 (2000).

- 3) M.Fischler and R.Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” Commun. ACM, vol.24, no.6, pp. 381–395, June 1981.
- 4) D.Nister, “An efficient solution to the five-point relative pose problem,” Computer Vision and Pattern Recognition (CVPR ’03), vol.2, pp. 195–202, 2003.
- 5) C.Tomasi and T.Kanade, “Shape and motion from image stream under orthography: A factorization method,” IJCV, vol.9, pp. 137–189, 1992.
- 6) A.Banno, K.Hasegawa and K.Ikeuchi, “Flying laser range sensor for scanning large-scale cultural heritages,” IPSJ SIG Notes CVIM, pp. 213–220, 3 2005.
- 7) R.Matsuhisa, H.Kawasaki, S.Ono, A.Banno and K.Ikeuchi, “Fundamental study of high accuracy estimation of vehicle motion using omni-directional image sequence,” The 7th ITS Symposium 2008, 2008.
- 8) M. Bertalmio, G. Sapiro, V. Caselles and C. Ballester, “Image inpainting”, Computer Graphics. SIGGRAPH, 2000.
- 9) Arturo Flores and Serge Belongie. “Removing pedestrians from google street view images”, In IEEE International Workshop on Mobile Vision, San Francisco, CA, June 2010.
- 10) 栗林宏輔, 子安大士, 前川仁, 小野晋太郎, 川崎洋, 池内克史, ” 平面群構造を利用した移