

EPI 解析を利用した画像統合による 都市空間のイメージベースレンダリング

小野 晋太郎^{*1} 川崎 洋^{*2} 影沢 政隆^{*1} 池内 克史^{*1} 坂内 正夫^{*1}

東京大学^{*1} 埼玉大学^{*2}

論文概要：エピポーラ平面画像 (EPI) を用いて複数ビデオカメラの画像を歪みなく統合する手法を提案する．一般にカメラの光学中心はカメラ内部にあり，それらを一点に集めることは物理的に困難である．従ってこれらの画像を統合する場合は，対象の三次元形状情報を用いない限り歪みが生じる．我々はこの問題に対し，時空間内で光学中心を一致させることで歪みなく画像を統合する手法を提案する．本手法では個々のカメラの撮影範囲にほとんどオーバーラップが必要なく，比較的自由にカメラを設置することが可能である．また，外部同期などの特別な装置を必要としないため，都市空間をはじめとした大規模な実空間のテクスチャ取得にも適している．

Image Based Rendering of Urban Spaces by Using EPI-Based Image Mosaicing

Shintaro Ono^{*1} Hiroshi Kawasaki^{*2} Masataka Kagesawa^{*1} Katsushi Ikeuchi^{*1} Masao Sakauchi^{*1}

The University of Tokyo^{*1} Saitama University^{*2}

Abstract: In this paper, we present an Epipolar Plane Image(EPI) - based multiple video camera calibration method, which enables simultaneous estimation of cameras' parameters and allows optical synchronization of video data. Using this method, in order to capture texture image of a large scale scene, users are only required to install multiple video cameras on top of the vehicle without setting exact configurations, and can take a video without using any external synchronization devices. Especially, the optical centers of cameras are made to coincide at one point in spatio-temporal space and therefore we can fuse these images into a single one without distortions.

Keyword: EPI, spatio-temporal image, IBR, video camera calibration, urban scene

1 はじめに

都市空間をはじめとした大規模な空間のモデル化や，そのための効率的なデータ取得方法はコンピュータビジョンにおける重要な研究テーマである．

これを実現する手法としては，主に対象となるオブジェクトの幾何情報と表面反射属性を利用して任意の視点からの画像を合成するジオメトリ (モデル) ベースの手法 (MBR) と，蓄積した画像を再構成することで新しい視点からの画像を生成するイメージベースによる手法 (IBR) が知られている．それぞれの手法においては品質

の高いデータを効率的に取得することが重要であり，特にあらゆる視点からのテクスチャを必要とする IBR においては極めて重要である．

効率的にデータを取得する手法としては，全方位カメラを用いて広い範囲を即座に撮影し，ビデオカメラで記録する手法が挙げられる．全方位カメラシステムには主に，一台のカメラと曲面鏡を組み合わせたものがあり，曲面鏡としては放物面 [1] や双曲面 [2, 3, 4]，円錐面 [5] 等が用いられている．しかしこれらの全方位カメラでは通常，全方位にわたるシーン全体を一つの画像として撮影するため，得られる画像の解像度が必然的に低くなる．

一方で、高い解像度の全方位画像を取得する手法として、複数のカメラを円形に配置し、それらの画像を統合する方法が提案されている [6]。また、さらに多くのカメラを用いた手法も提案されている [7]。しかし、このようなカメラシステムではそれぞれのカメラの光学中心を一致させることが物理的に困難であるため、統合した画像に歪みが生じる。さらに、個々のカメラの同期をとることも容易ではない。

そこで本論文では、光学中心の不一致による歪みを伴わず、カメラ間の同期をとる必要のない簡便なキャリブレーション手法を提案する。提案する手法は時空間画像解析 (EPI) に基づいているため、安定した処理が可能である。

本稿の構成は以下のとおりである。2 章では本手法の特徴を関連研究とあわせて述べ、3 章でアルゴリズムの詳細について述べる。4 章では歪みに関する誤差評価とカメラの配置に関する考察を行う。5 章では実際に屋外空間において実験を行った結果を示す。6 章でまとめとする。

2 手法の概要と関連研究

本研究の目的は、複数のビデオカメラから構成される全方位カメラを用いて都市空間をはじめとした大規模な環境を取得し、歪みなく画像を統合することである。これを実現するためには、複数ビデオカメラのキャリブレーションが不可欠である。

2.1 複数ビデオカメラのキャリブレーション

複数カメラのキャリブレーション、すなわちカメラ間のパラメータ推定はコンピュータビジョンの分野において基本的かつ重要な研究課題である。複数のカメラで撮影された画像間の対応関係が求まれば、各種のキャリブレーションアルゴリズムを適応することが可能である。しかしながら、ビデオ映像などのように密にサンプリングされた対象に対して、これらのアルゴリズムを適応することは容易ではない。また対応点がそれぞれのカメラ画像に明確に写り込んでいる必要があるため、撮影範囲が比較的大きくオーバーラップしていなければならない。

動画像の解析手法として、カメラの動きを制限することで、安定に撮影対象の三次元形状を推定することが可能な EPI 解析 [8] が知られている。EPI 上では画像列中にある特徴点の軌跡が直線となって現れるため、特徴点

のトラッキングをエッジ検出に帰着することができる。エッジ検出処理は通常、ハフ変換により精度良く行うことができる。

これまで EPI 解析は、主として対象の三次元形状情報を獲得するために用いられてきた。しかし前述の通り、EPI 上の直線は時系列画像中の特徴点の軌跡を表しているため、複数のカメラ間でこの直線どうしの対応関係を得ることが出来ればキャリブレーションを行うことが可能である。また、EPI を用いた手法ではオーバーラップがほとんど必要ないため、広い領域を効率よくキャプチャすることができる。

2.2 大規模シーンのモデル化

都市空間をはじめとした大規模シーンのモデル化に関する研究はこれまでに数多く行われてきた。特に、これまでジオメトリベースによる手法が多数提案されている。最近では、Frueh らが車載したレンジセンサを用いてテクスチャ付きの三次元モデルを効率的に取得・構築する手法を提案している [9]。

一方、イメージベースにより大規模空間を構築する試みも行われてきた [6, 10] が、そのデータ取得の困難さから、これまでは盛んに行われてはこなかった。そこで本稿では、IBR に適用できる効率的なテクスチャ取得を目指す。

3 時空間における複数カメラのキャリブレーション

本研究では EPI を利用して複数カメラを時空間内においてキャリブレーションすることで、歪みのないパノラマ画像を合成する。以下では、その原理について述べ、続いて EPI を用いてこれを実現する手法について説明する。

3.1 時空間における光学的中心の一致

既に述べた通り、複数カメラの光学中心を一致させることは通常、物理的に困難である。しかし、時空間、すなわち時間差とカメラ位置の差を考慮することで容易にこれを実現することができる。

3.1.1 原理

図 1(a) に示すように、カメラ 1 が時刻 T_1 において地点 $P_0 = (x_0, y_0, z_0)$ から画像を撮影するものとする。時刻 T_2 においては、カメラ 2 が同じく地点 P_0 から画像を撮影する (図 1(b))。また同様に、時刻 T_3 においては、カメラ 3 が地点 P_0 から画像を撮影する (図 1(c))。この

ように、時刻は異なるが、すべてのカメラが同じ地点 P_0 から、それぞれの方向の画像を撮影したことになる。これを実現するためには、各々のカメラが同じ地点を通過する必要がある。

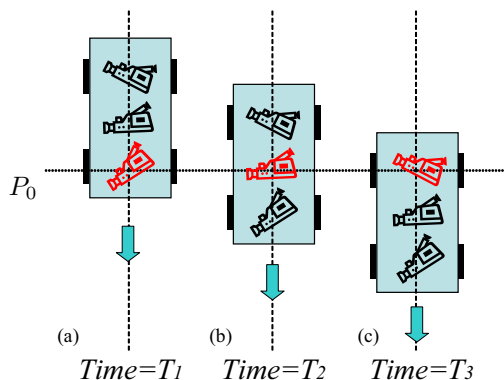


図 1: 光学的な同期

したがって、カメラは図 2(a)(b) に示すように、移動方向に対して平行な直線上に並べて配置するのが理想的である。無論、カメラの配置間隔や方向は未知のまま任意に設定することができる。このように一直線上にカメラが配置できない場合は、完全に歪みのない統合画像を得ることは不可能である。ただし、要求される精度や目的によっては、図 2(c) のように奥行きをずらして配置することも可能である。カメラおよび対象物体の位置と統合画像における歪みの大きさに関する評価については後に詳しく述べる。

個々のカメラが撮影する範囲のオーバーラップは、統合後の画像に隙間が生じない程度にあれば充分である。本研究で提案する EPI を利用したキャリブレーション手法では、たとえ空間的なオーバーラップがまったく存在しない場合でも処理が可能である。

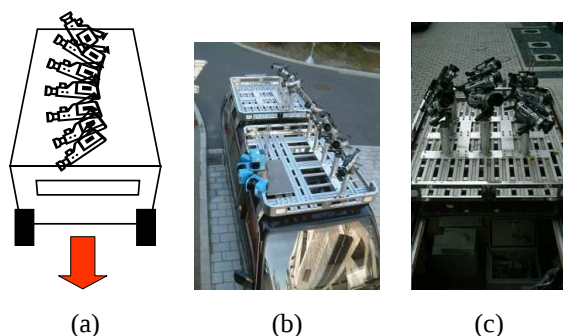


図 2: 複数カメラの配置

3.2 EPI を利用したキャリブレーション

ここでは、時空間におけるカメラキャリブレーションを EPI 解析を用いることで効率的に実現する。カメラは任意の方向を向いているため、EPI を生成するにあたってまず平行化処理を行って向きに関する正規化を行う必要がある。以下ではまず平行化処理について述べ、次に EPI 解析によるキャリブレーション手法の詳細について述べる。

3.2.1 平行化

都市空間における建物やその他のオブジェクトは地面に対して水平・垂直な二種類の直線で構成されていることが多い。また、平行化処理に必要なパラメータは、カメラの向きが変化しない限り、全体を通して一度だけ求めれば充分である。そこで、直線を選択は人手により行い、それらの交点の座標を求める。この交点は消失点と呼ばれる。理論的には消失点に向かう直線はすべて一点で交わるが、実際にはノイズや解像度の問題から一点で交わることはない。そこでこれらの交点の平均値を取り、これを求める交点とした。

消失点を利用すると、回転行列と焦点距離は以下のように求められる。まず r_1 と r_2 を以下のように定義する。

$$r_i = \begin{pmatrix} u_i \\ v_i \\ f \end{pmatrix} \quad (i = 1, 2) \quad (1)$$

ここで (u_i, v_i) は消失点の画像座標、 f はカメラの焦点距離（ピクセル値）である。これらのベクトルは実世界中での水平・垂直方向にそれぞれ平行である [10]。よって r_1, r_2 は直交し、内積は 0 となる。したがって f は以下のように求められる。

$$f = \sqrt{-(u_1, v_1) \cdot (u_2, v_2)} \quad (2)$$

r_1, r_2 を正規化して $\hat{r}_1, \hat{r}_2$ とし r_3 を

$$r_3 = \hat{r}_1 \times \hat{r}_2 \quad (3)$$

とすると、回転行列 R は以下ようになる。

$$R = \begin{pmatrix} r_1 & r_2 & r_3 \end{pmatrix} \quad (4)$$

回転行列 R と焦点距離 f が得られれば、射影変換によって平行化を行うことができる。実際に平行化を行った例を図 3 に示す。

3.2.2 EPI どうしのマッチング

EPI とは、図 4 のように動画像をフレーム毎に積み重ね、時間軸に沿って水平に切断した断面に現れる時空間



図 3: 並行化を行った例

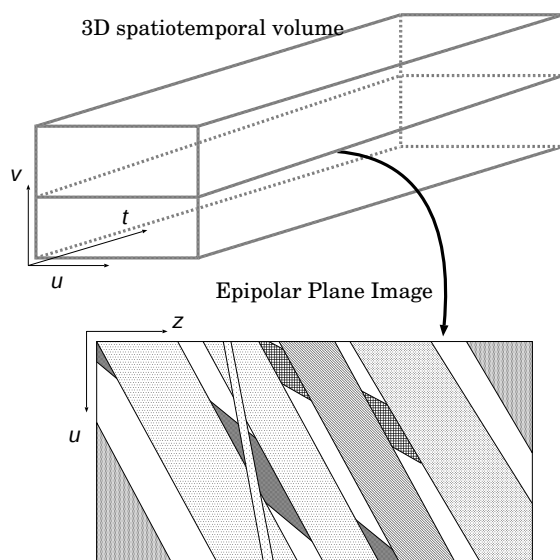


図 4: EPI と座標系の定義

画像である。カメラが等速で移動している場合、移動方向に対して光軸が直角に向けられているカメラから得られる EPI では対象の特徴点が描く軌跡が直線のエッジとなって現れ、このエッジの傾きがその点の奥行きを表している。その他の方向を向いたカメラでは EPI 上のエッジは曲線となるが、平行化処理を施して向きを正規化した画像からは直線のエッジが得られる。

それぞれのカメラから得られた EPI どうしをマッチングさせることで、時空間において光学中心を一致させることができる。これを図 5 を用いて説明する。

いま、図 5(i) において、2 台のカメラ a, b からそれぞれの EPI が同じ画像平面上に得られているとする。これを EPI 間の水平移動のみで 1 次元のマッチングを行い、重ね合わせた例が図 5(ii) である。この場合は時空間を考慮しておらず、同じ時刻に異なる地点から得られた画像フレームを、適当なデプスを仮定した上で射影変換などの手法により統合した場合に相当する。図中 (ii) は、カメラ手前の壁面 (物体 No.1, 2) にデプスを仮定してマッチングを行った結果である。実際には、それ以外

の対象物 (物体 No.3) はデプスが仮定したものと異なるため、統合結果に歪みを生じる。また、このようなマッチングを行う場合は、必然的にカメラ a, b の撮映範囲に十分なオーバーラップが必要となる。

次に、図 5(iii) は二次元 EPI マッチングを行った結果である。これは、2 つのカメラ映像に対して時間と空間を同時にずらしながら最も適合する点を探索することに相当する。この場合、EPI 上の全ての直線が滑らかにつながっている。これはマッチングの結果、時空間内において光学中心が一致していることに他ならない。この二次元 EPI マッチングの結果を利用すれば、歪みなく画像を統合することができる。このときの EPI 画像の移動量 Δz が、カメラ a, b 間の間隔に相当する。また、2 台のカメラの撮影範囲が空間的にまったくオーバーラップしていない場合でも、このマッチングでは時間軸方向にも探索を行うため、撮影対象が時間的にオーバーラップしていれば安定にマッチングを行うことができる。

図 6 に 3 台のカメラ画像に平行化を施したものから得られた EPI と、実際に二次元 EPI マッチングを行った結果を示す。複数の EPI が滑らかに連結されていることがわかる。

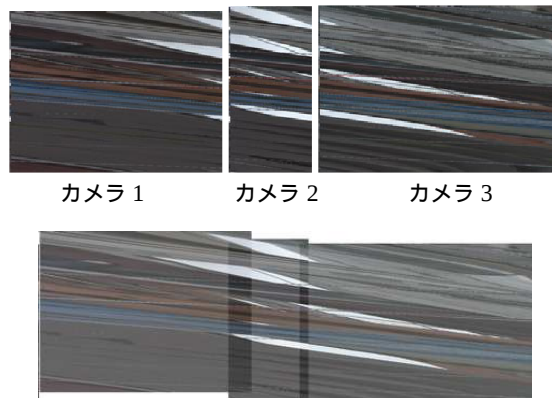


図 6: 再スケールした EPI とマッチングの結果

4 統合画像における歪み誤差の評価

EPI によるキャリブレーションを実現するためには、それぞれのカメラから得られた EPI がすべて同じ画像平面上に投影されている必要がある。すなわちカメラは、図 2(a) のようにすべて進行方向に対して平行に、一直線上に配置するのが理想的であると言える。しかし、そのような配置はカメラの台数にも設置上の制約から限界があり、必ずしも一般的な方法とは言えないことがある。

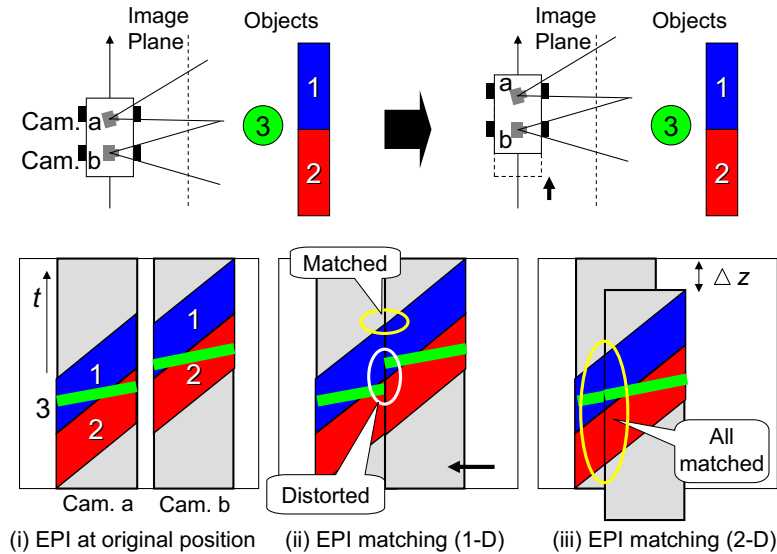


図 5: EPI どうしのマッチング

本章では、通常の射影変換を行った場合にカメラの設置位置が統合画像に及ぼす歪みの大きさについて解析し、一般的なカメラ配置について考察する。

4.1 歪み量の解析

図 7 のように、基準となるカメラを原点とし、カメラの進行方向を x 軸にとるような座標系を考え、他のカメラが $(\Delta x, \Delta y)$ に設置されているとする。このとき、デプスを $x = D$ の位置に仮定して射影変換により画像統合を行うと、実際のデプスが $x = d$ である点に対して生じる水平方向の歪み量は、図中の diff に相当し、

$$\text{diff} \propto \frac{\Delta y - \Delta x \tan \theta}{D} \frac{d - D}{d - \Delta x} \quad (5)$$

と表される。カメラを図 7 の (A), (B) のようにずらして設置した場合の歪み量はそれぞれ、図 8(A), (B) のようになる。

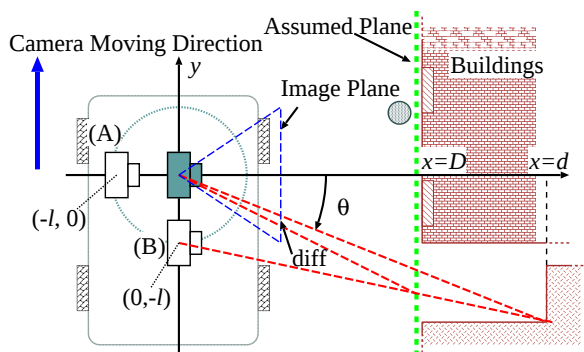


図 7: カメラ配置による歪み量

これらのグラフから、歪み量はデプス値 d および仮定するデプス値 D によって大きく変化することがわかる。仮定デプス値が実際のデプス値に等しいときは歪み量は 0 となる。

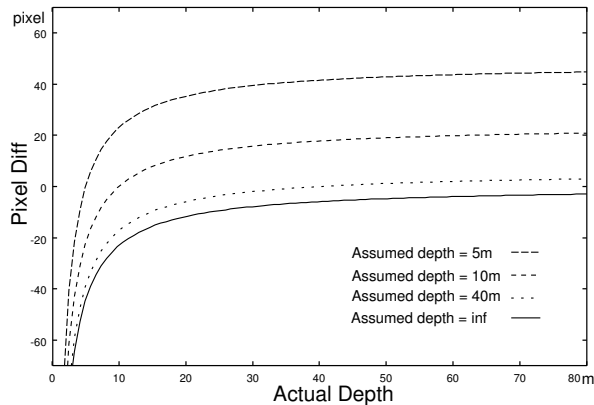
4.2 一般的なカメラ配置

前節で解析した歪み量から一般的なカメラ配置について考察する。

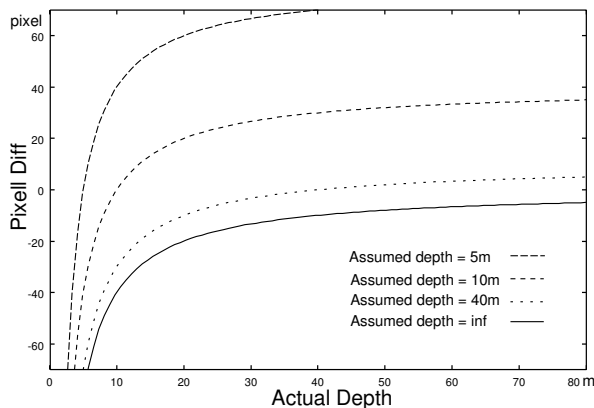
図 8 を見ると、図 7 (B) のように進行方向に対して平行に配置したカメラでは、射影変換により画像を統合すると、比較的大きな歪みが生じる。さらに、歪みの変化率も実際のデプス および 仮定するデプスのずれに対して大きく変化することがわかる。しかしながら、これらの歪みはすべて、前章に述べたように時空間において光学中心を一致させることですべて除去することが可能である。

次に、図 7 (A) のようにデプス方向にずらして配置したカメラにおいて生じる歪みについて考察する。このようにカメラを設置した場合は時空間においても光学中心を一致させることができず、歪みを完全になくすことは必然的に不可能である。図 8 より、この歪みには以下のような特徴があることがわかる。

- 進行方向のずれから生じる歪み (グラフ (B)) よりも比較的小さい
- 極端に大きな歪みが生じるのは、対象物の実際のデプスが小さく、カメラから近い場合に限られる
- 仮定するデプスが実際のデプスに近いときは、歪み



(A)



(B)

図 8: 歪み量の評価

の変化率も小さい

したがって、対象空間のデプスが概ね一定の位置に仮定できる場合や、カメラから近い位置にある対象物が無視できるような場合は、要求する品質によっては図 7 (A) のようにずらして配置した場合でも実用的であると言える。

都市空間はこの条件に合致することの多い一例である。最も支配的な対象物体となるのは道路に沿って並んだ建物であることが多く、比較的カメラの近くに存在する電柱などは、目的によっては無視できることも多い。

本研究では、図 2 (b) のように進行方向に沿って 9 台 × 1 列のような配置のほか、図 2 (c) のように 3 台 × 3 列、あるいは 3 台 × 2 列のように配置して実験を行った。

5 実験

5.1 画像の取得プロセス

計測車両の屋根に複数のビデオカメラを搭載し、道路を走行しながら進行方向に対して右側の撮影を行った。今回の実験では、計測車両は等速直線運動をするものと仮定した。

等速直線運動が仮定できない場合は、撮影と同時に車速センサ、GPS、ジャイロセンサなどの記録を同時に行い、走行経路を直線区間ごとに区切って処理を行う。また、車速センサや画像処理等により推定した実際の車速から、等速走行で撮影した場合のように映像を正規化する。

5.2 実験 (I): 複数カメラ画像の統合

ここでは提案手法によって複数枚の画像をつなぎ合わせてパノラマ画像を生成する。まず、図 2(c) のように、進行方向に対して平行に 3 台、デプス方向に 2 台配置して撮影した画像の統合を行った。図 9 (a)-(f) はオリジナル画像であり、図 9 (g) は合成したパノラマ画像である。この合成画像では歪みはほとんど見られない。これは建物がほぼ同じ平面上にあるため、射影変換のみでも歪みのない画像統合が可能なたためである。また、図 10 に別の合成例を示す。この図では電柱付近に歪みがみられる。これは電柱が他の建物に対して著しくカメラに近い位置にあるため、図 2(c) のようにカメラが配置された場合、歪みが無視できなくなるためである。

これに対して、図 11 は、図 2(b) のように、理想的なカメラ配置により撮影した画像を統合した例である。光学中心が完全に一致しているため、後方の建物も前方の電柱も、歪み無く統合できていることがわかる。

5.3 実験 (II): 自由視点画像生成

この実験では、得られた統合画像データを用いて Takahashi らの IBR 手法 [12, 10] により、走行経路上以外の新しい視点からの映像を合成した。

図 12(d) に走行経路と対象環境の位置関係を示す。図 12(a)-(c) は、(d) に示した仮想的な視点から見た場合の合成画像である。いずれの合成画像においても、奥行き位置にかかわらず歪みなく画像を統合できている。また、仮想視点が移動するに伴って図中の木が正しい位置に描かれ、背後の建物を適切に遮蔽していることがわかる。

また、図 12(e)(f) は仮想視点を (a) と同じ位置に置き、1 台および 3 台のカメラから合成した画像である。複数



図 10: 6 台カメラ画像の統合結果：カメラは図 2(c) のように 3 台 × 2 列に配置した．カメラから近い位置にある電柱には歪みが生じている．



図 11: 9 台カメラ画像の統合結果：カメラは図 2(b) のように 9 台 × 1 列に配置した．デプスによらず歪みなく統合することができる．

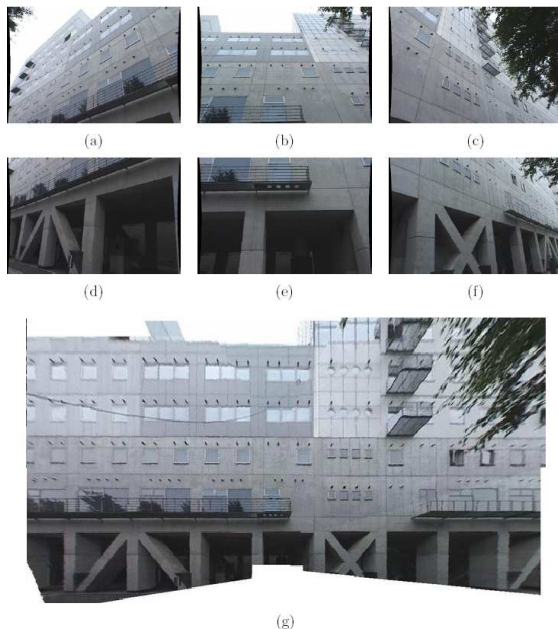


図 9: (a)-(f): 各カメラの撮影画像 (g): 6 台カメラ画像の合成結果

台のカメラを用いることで視野が広がり，IBR 手法を用いた場合でも死角なく，自然な画像合成ができていることが分かる．

6 結論

本論文では，EPI による複数カメラのキャリブレーション手法を提案した．本手法によれば，複数のビデオカメラを計測車両の屋根に設置し，走行しながら撮影するだけで効率よく大規模シーンの画像を取得し，歪みなく統合することができる．

一般に，カメラの光学中心はカメラの内部にあるため，複数カメラの光学中心を一点に集めることは物理的に困難である．これに対して我々は，移動しながら対象を撮影する特徴を利用し，時空間，すなわち時間差と空間上のカメラの位置の差を同時に考慮することによって複数カメラの光学中心を一致させることに成功した．

また，本手法の有効性を確認するため，実空間を対象として実験を行った結果，あらゆるデプスにおいて歪みのない画像を生成することに成功した．また，全方位画像の特性を利用して，撮影地点以外からの視点によるパノラマ画像を合成することができた．

参考文献

- [1] Shree K. Nayar, "Omnidirectional video camera," Proc. of Computer Vision and Pattern Recognition, 1997.
- [2] Y. Onoue, K. Yamasawa, H. Takemura and N. Yokoya, "Telepresence by realtime view-dependent image generation from omnidirectional video streams," Computer Vision and Image Understanding, vol.71, no.2, pp.154-165, Aug. 1998.
- [3] Y. Yagi and S. Kawato, "Panorama scene analysis with conic projection," Proc. Int. Conf. on Intelligent Robots and Systems, pp.181-187, Jul. 1990.
- [4] K. Yamasawa, Y. Yagi and M. Yachida, "New realtime omnidirectional image sensor with hyperboloidal mirror," Proc. 8th Scandinavian Conf. on Image Analysis, pp.1381-1387, May 1993.
- [5] J.Y.Zheng and S.Tsuji, "Panoramic representation of scenes for route understanding," Int. Conf. on Pattern Recognition, pp.161-167, June 1990.
- [6] M. Hirose and E. Takaaki, "Building a virtual world from the real world," Proc. of International Symposium on Mixed Reality, pp. 183-197, Mar. 1999.
- [7] 遠藤隆明, 谷川智洋, 広田光一, 廣瀬通孝「超多眼カメラによる全天周画像の再構成」情報処理学会論文誌, Vol.43, No.SIG 11(CVIM 5), p.43-52 (2002.12)
- [8] R. Bolles, H. Baker and D. Marimont, "Epipolar plane image analysis: an approach to determining structure from motion," Int. J. of Computer Vision, vol.1, pp.7-55, 1987.
- [9] C. Fruh and A. Zakhor, "3D model generation for cities using aerial photographs and ground level laser scans," Computer Vision and Pattern Recognition, vol. 2, pp. 31-38, 2001.
- [10] T. Takahashi, H. Kawasaki, K. Ikeuchi and M. Sakauchi, "Arbitrary view position and direction rendering for large-scale scenes," Computer Vision and Pattern Recognition, vol. 2, pp. 296-303, June 2000.
- [11] S. D. Ma, "A self-calibration technique for active visio systems," IEEE Trans. RA, vol. 10, pp. 7-25, 1993.
- [12] H.Y.Shum and Li Wei He, "Rendering with concentric mosaics," ACM SIGGRAPH, pp. 299-306, 1999.

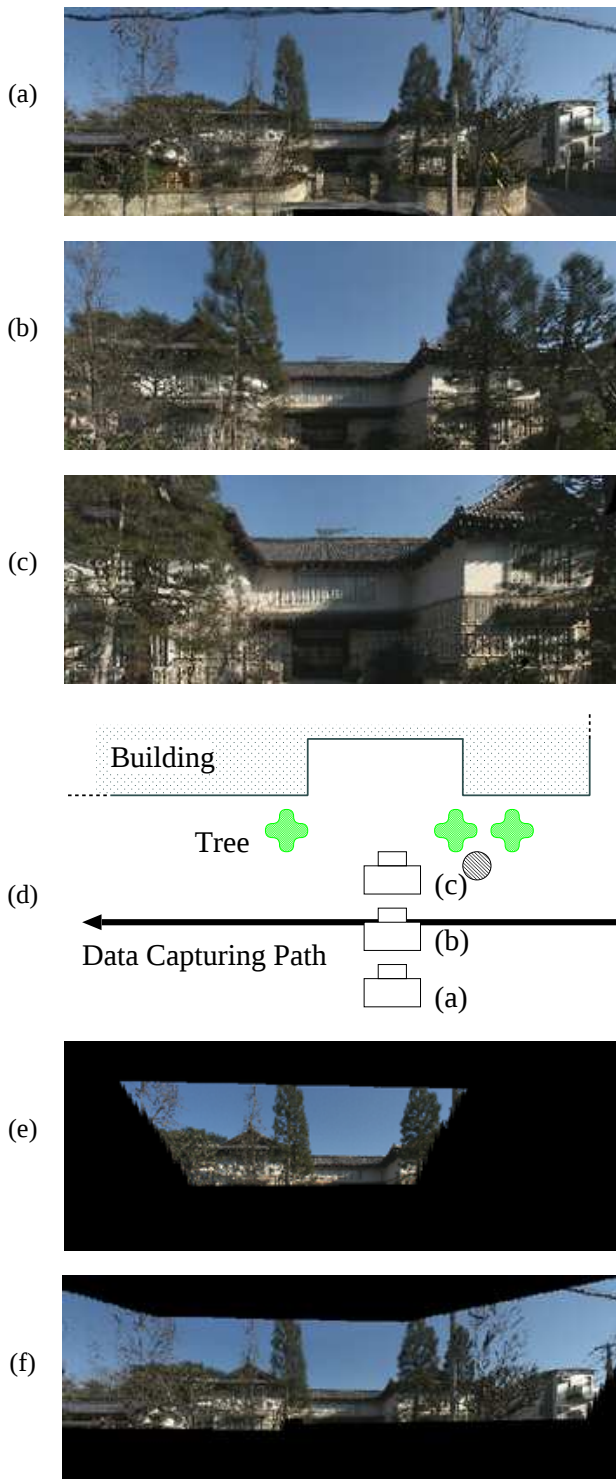


図 12: (a)-(c):自由視点画像の生成 (d):カメラ位置 (e):カメラが 1 台の場合 (f):カメラが 3 台の場合