

EPI解析による複数ビデオカメラの画像統合

小野 晋太郎 川崎 洋 池内 克史 坂内 正夫

東京大学生産技術研究所

〒 153-8505 東京都目黒区駒場 4-6-1 Tel : 03-3452-7885

E-mail: {onoshin, ki}@cvl.iis.u-tokyo.ac.jp, {h-kawa, sakauchi}@sak.iis.u-tokyo.ac.jp

あらまし

本論文では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無く統合する手法を提案する。複数台のビデオカメラを車両の上に設置し、走りながら撮影することで都市空間をはじめとした大規模なシーンの映像を効率的に取得することが出来る。しかし、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が物理的に難しく、このため、これら複数の画像を、三次元形状情報を用いることなく統合すれば、統合した画像に歪みが生じる。我々はこの歪みを除去する方法として、空間的に光学中心を一致させることが難しい場合であっても、これを時空間内で実現する手法を提案する。

また、提案する手法では、複数のビデオカメラを比較的自由に設置することが可能であり、外部同期などの装置を必要としないため、実際の都市空間におけるテクスチャ取得等に適している。

キーワード 全方位画像, 時空間画像解析, 画像合成, 映像データベース

Image Mosaicing of Multiple Video Camera using EPI Analysis

Shintaro ONO Hiroshi KAWASAKI Katsushi IKEUCHI Masao SAKAUCHI

Institute of Industrial Science, The University of Tokyo

4-6-1 Komaba, Meguroku-ku, Tokyo JAPAN 153-8505 Tel : 03-3452-7885

E-mail: {onoshin, ki}@cvl.iis.u-tokyo.ac.jp, {h-kawa, sakauchi}@sak.iis.u-tokyo.ac.jp

Abstract

In this paper, we present an Epipolar Plane Image (EPI)- based multiple video camera calibration method which enables simultaneous estimation of the multiple video cameras' parameters and allows optical synchronization of video data. Using our proposed method, in order to capture a large scale scene's texture image, a user is only required to install multiple video cameras on top of his/her car without setting exact configurations, and can take a video without using a special external synchronization device.

Since the optical centers of cameras are usually inside the cameras, and bringing them together is usually impossible, distortions inevitably occur in fused images. However, with our proposed method, the optical centers of the multiple cameras images are made to coincide at one point in spatio-temporal space and thus we can fuse these images into a single one with no distortions.

key words Omni-directional camera, Spatio-temporal analysis, Image Synthesis

1 まえがき

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、対象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、品質の高いデータの効率的な取得が極めて重要であるが、あらゆる視点からのテクスチャを必要とする IBR にとっては、特に重用である。

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、これをビデオカメラで記録を行うことが考えられる。効率的な全方位カメラシステムとして、一台のカメラと曲面鏡を用いたものがある。曲面鏡としては放物面 [1] や双曲面 [2, 3, 4]、円錐面 [5] 等がある。しかしこれらの全方位カメラでは通常全方位のシーン全体を一枚の画像に撮影するため、得られる画像の解像度は必然的に低い。

一方で、全方位画像を高い解像度で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ある [6]。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難しいため統合した画像が歪む。さらに、個々のカメラの同期をとることも簡単ではない。

そこで、本稿では光学中心の不一致による歪みを伴わ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 章では実際に屋外空間において実験を行った結果を示す。5 章でまとめる。

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する大規模な環境をコンピュータ内に取得することである。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出来る。

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簡単ではない。

ビデオ解析手法として、カメラの動きを制限することで、ロバストに形状を推定可能な EPI 解析 [7] が知られている。EPI 上では画像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、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来た。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、この直線どうしの対応関係を得ることが出来れば、複数カメラのキャリブレーションを実現することが可能である。

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も数多く行われてきた。

特に、これまでモデルベースによる手法が多数提案されている。最近では、柴崎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 [8]。

一方、IBR により構築する試みも行われて来た [6]、[9]。しかしながら、IBR による大規模空間の構築は、そのデータ取得の困難さから、これまで盛んに行われては来なかった。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 時空間での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 によりそれを実現する手法について述べる。

3.1 時空間での光学的中心の一致

既に述べた通り、通常、複数カメラの光学中心を一致させることは物理的に難しい。しかし、時空間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-(i) に示すように、カメラ # 1 が時刻 T_1 において地点 (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 (x_0, y_0, z_0) から画像を撮影する (図 1-(ii))。また同様に、時刻 T_3 においては、カメラ # 3 が同じく地点から画像を撮影する (図 1-(iii))。このように、時刻は異なるが、すべてのカメラが同じ地点 (x_0, y_0, z_0) から、それぞれの方向の画像を撮影したことになる。

勿論、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。

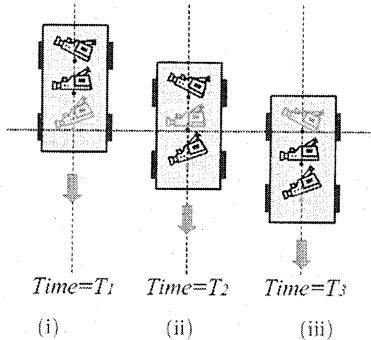


図 1: 光学的な同期

3.1.2 カメラの配置

従って、カメラは移動方向に対して平行な直線上に並べて配置されていることが望ましい (図 2-(a))。本研究では、図 2-(b) のように、1 列となっているものも、図 2-(c) のように、3 列となっているものも制作し実験を行った。以下に、設置条件を述べる。

- 対象領域がカメラの撮影範囲に収まっている必要がある (ただし、必ずしも画像同士がオーバーラップしている必要はない)。

- カメラ中心は 図 2-(a) のように移動方向と平行に直線に沿って並べることが望ましい。

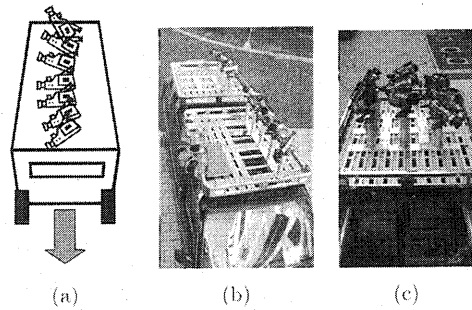


図 2: 複数カメラの設置 (a:理想的配置, b,c:実際の配置)

3.2 EPIによるキャリブレーション

ここでは、時空間での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できることを示す。

カメラは任意の方向を向いているため、EPI を生成するためには平行化処理が必要である。そこで、以下ではまず平行化処理について述べる。

また、EPI によるキャリブレーションを実現するためには、全ての EPI が同じ画像平面上に投影されている必要があるが、カメラは必ずしも図 2-(a) のように理想的に配置されている訳ではない。そこで、続いて、それぞれのカメラの進行方向に対して垂直な方向のずれ (デプス) を求める手法を述べる。

最後に EPI によるキャリブレーション手法について述べる。

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ることが多い。また、平行化のパラメータは全体を通して一度だけ求めれば良い。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。対象画像中では、水平方向・垂直方向の二つ消失点が存在する。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

ずや解像度の問題からそのようにはならないため、これらの交点の平均値をとり、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、 \hat{r}_3 を

$$\hat{r}_3 = \hat{r}_1 \times \hat{r}_2 \quad (3)$$

とすると、回転行列 R は以下ようになる。

$$R = \begin{pmatrix} \hat{r}_1 & \hat{r}_2 & \hat{r}_3 \end{pmatrix} \quad (4)$$

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。図 3 に平行化を行った例を示す。画像の平行化に成功していることがわかる。

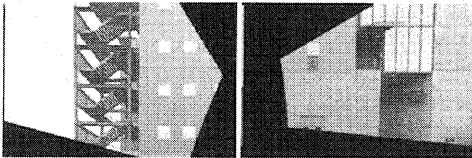


図 3: 並行化を行った例

3.2.2 デプスの推定

平行化後の画像上の水平線を時間方向に積み重ねることによって EPI を作成できる。本研究では複数台のビデオカメラを用いるため、各々のカメラからそれぞれの EPI が作成される。

EPI から検出されたエッジは以下のように表せる。

$$z = \frac{d_n}{vf_n} \cdot u \quad (5)$$

図 4 に $u-z$ 座標系の定義を示す。

n はカメラの番号、 v はカメラの移動速度、 d_n はカメラ n から物体までの奥行き値である。これはカメラの位置によって変化し、 $d_n = d_{world} - y_n$ と表される。

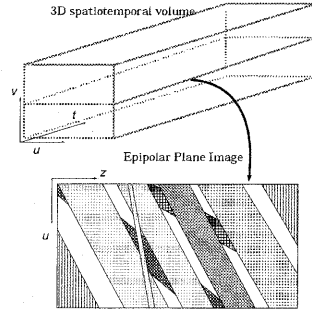


図 4: EPI と座標系の定義

したがって

$$y_n = d_{world} - d_n \quad (6)$$

$$= d_{world} - a_n \cdot vf_n \quad (7)$$

ここでエッジの傾き $a_n = \frac{z}{u}$ は EPI から求められる。よって t_y は以下のように求められる。

$$t_y = y_m - y_n = (a_m f_m - a_n f_n) \cdot v \quad (8)$$

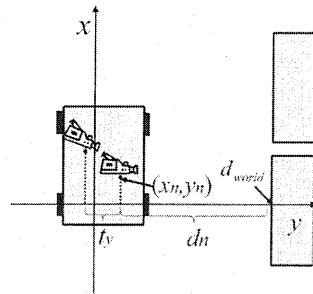


図 5: ワールド座標系におけるカメラ位置

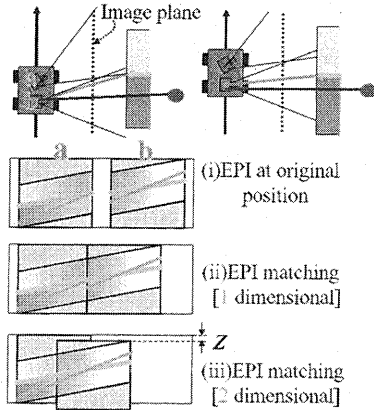


図 6: EPI どうしのマッチング

3.2.3 EPI どうしのマッチング

EPI どうしをマッチングさせることで、時空間での光学中心の一致を実現することが出来る。これを図 6 を用いて説明する。ここでは、解析を簡単にするため、あらかじめ外部の装置によって映像データの同期がとれているものとする。

今、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 (図 6-(i))。

これを EPI 間の 1 次元マッチングにより重ね合わせた例が図 6-(ii) である。これは、時空間を考慮せずに、画像をホモグラフィなどの手法により統合した場合に相当する。物体の実際のデプスは仮定したデプスと異なるため、画像を歪み無く統合することは出来ない。

図 6-(iii) が二次元 EPI マッチングを行った結果である。この場合、EPI 上の全ての直線がなめらかにつながっている。これは、マッチング結果 z が、時空間における光学中心の一致を実現しているためである。この二次元 EPI マッチングの結果 z と u を利用すれば、直接歪みのない画像の合成が実現できる。

図 8 に実際に二次元 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。

4 実験

本手法の有効性を確かめるため、実際に屋外で実験を行った。

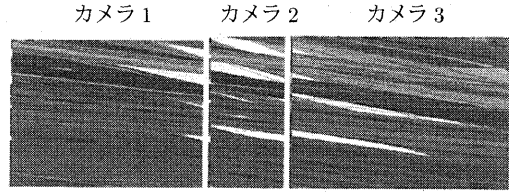


図 7: 再スケールした EPI

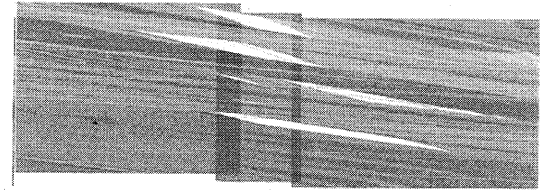


図 8: EPI マッチングの結果

4.1 画像の取得プロセス

計測車両に複数のカメラを搭載し、道路を走行しながら撮影を行った。また、撮影と同時に車速センサや GPS、ジャイロセンサなどの記録も同時に行う。画像はあらかじめ画像処理によりレンズ歪みなどが除去してある。

4.1.1 走行経路

EPI は等速直線運動を仮定しているが、実際にこの条件で計測することは出来ない。そのため、本実験では、まず GPS やジャイロスコープなどのセンサを用いて、ビデオ映像を直線区間ごとに分割する。

続いて、車速センサや、画像処理等により推定した実際の車速により、あたかも等速で撮影したかのような映像にビデオを調整する。

4.2 実験 (I): 複数カメラの統合

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) のように配置して撮影した複数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。

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、理想に近い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分かる。

4.3 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、IBR 手法 [11][9] により新しい視点からの画像を合成した。

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。

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、窓が建物によって正しく遮蔽されていることがわかる。

また、図 12(f) は仮想視点を (a) と同じ位置に置き、3 台のカメラから合成した画像である。複数台のカメラを用いることで視野が広がり、IBR 手法を用いても、死角がなくなり、自然な画像合成が出来ていることが分かる。

5 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。本手法を用いれば、大規模シーンの画像を取得するためには複数のビデオカメラを計測車両の天井に設置し、自由に走行しながら撮影するだけで良い。

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難しい。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、時間内において光学中心の一致を実現した。

本手法の有効性を確認するため、実シーンにおいて実験を行った結果、歪みない新しい視点からのパノラマ画像を合成することができた。

参考文献

[1] S. K. Nayar, "Omnidirectional video camera," Proceedings of Computer Vision and Pattern Recognition, 1997.

[2] Y. Onoue, K. Yamasawa, H. Takemura and N. Yokoya, "Telepresence by real-time 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, July 1990.

[4] K. Yamasawa, Y. Yagi and M. Yachida, "New real-time 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," International Conference on Pattern Recognition, pp. 161-167, June 1990.

[6] M. Hirose and E. Takaaki, "Building a virtual world from the real world," Proceedings of International Symposium on Mixed Reality, pp. 183-197, March 1999.

[7] 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.

[8] 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.

[9] T. Takahashi, H. Kawasaki, K. Ikenuchi 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.

[10] S. D. Ma, "A self-calibration technique for active vision systems," IEEE Trans. RA, vol. 10, pp. 7-25, 1993.

[11] H.-Y. Shum and Li-Wei-He, "Rendering with concentric mosaics," ACM SIGGRAPH, pp. 299-306, 1999.

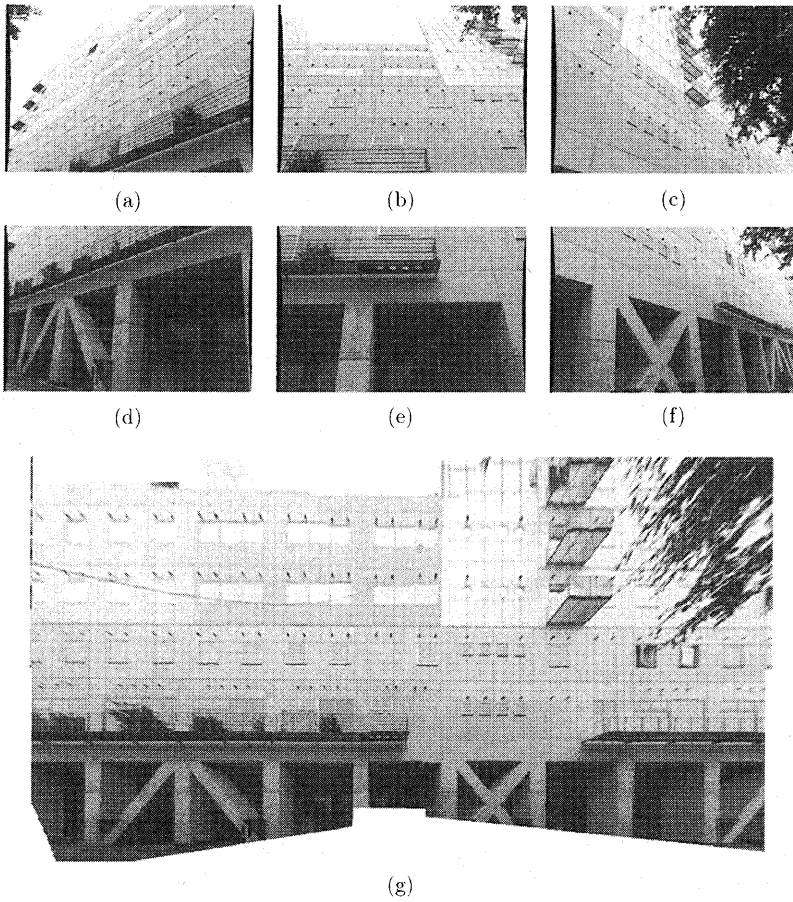


図 9: a-f: 各カメラの撮影画像 (レンズ歪み除去後), g: 6 台カメラ画像の合成結果 1

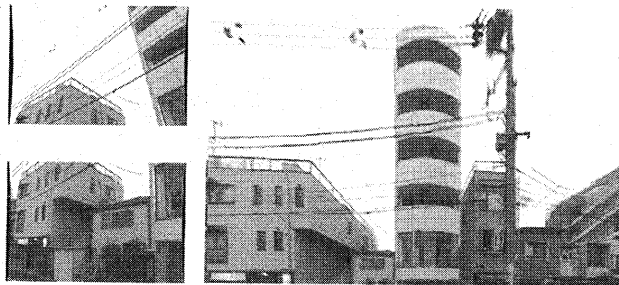


図 10: 6 台カメラ画像の合成結果 2



図 11: 9 台カメラ画像の合成結果

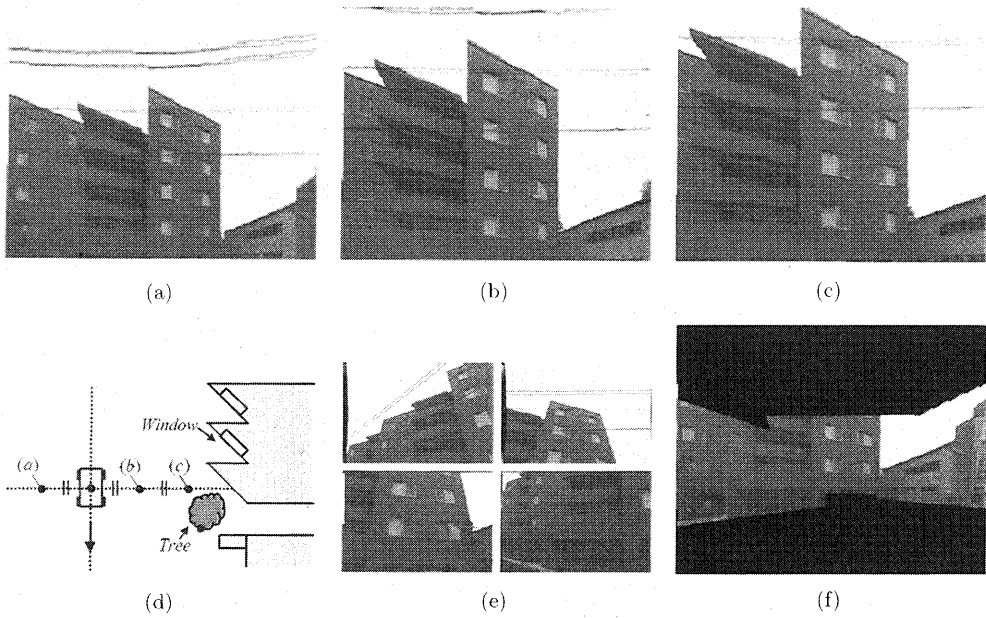


図 12: (a)(b)(c): 本手法により合成した画像 (d): カメラ位置 (e): 複数カメラからの入力画像 (f): カメラを 3 台だけ利用した場合