

動画画像解析による仮想化都市空間の構築

全 炳東†, 納富幹人†, 小澤史朗†

†東京商船大学 (東京都江東区越中島 2-1-6)
zen@ipc.tosho-u.ac.jp

あらまし この報告では、動画画像解析による都市景観の3次元モデル化手法の提案と、その応用・展開に関して述べる。都市の3次元モデル化は、都市工学、交通工学、教育、アミューズメントなどの広い分野で活用される重要な情報基盤であるが、その入力・構築は基本的に手作業に依存している。そのため、広範囲な空間のモデル化には莫大なコストが必要であり、用途が限定されたり、対象物を限定せざるを得ないのが現状である。この研究では、EPI解析による都市景観の3次元情報獲得の効率化と、数値地図を核とする地理情報システム上での、異種センサ情報との統合を目指している。

キーワード 動きステレオ, エピポーラ平面画像解析, 数値地図

Virtual City Space: A Construction of 3D City Scene through Moving Image Analysis

Heitou ZEN†, Mikito NOUTOMI†, Shiroh OZAWA†

†Tokyo University of Mercantile Marine
2-1-6, Etchujima, Koto, Tokyo 135
zen@ipc.tosho-u.ac.jp

Abstract A framework for 3D modeling of an urban scene is discussed in this report. The 3D geometric and texture information of urban scene is one of most expecting in many area, such as civil engineering, traffic engineering, education, amusement *etc.* The enhanced EPI (Epipolar Image Analysis) is employed to get these info. as a motion stereo method in our work. Putting these info. into the existing GIS (Geographic Information System) gives rise to more realistic looks and quick conformability of the system.

key words Motion stereo, EPI analysis, Digital map

1 はじめに

都市景観の3次元モデルは、都市工学、道路交通、教育、マーケティング、アミューズメントなど、広い分野で活用される重要な情報基盤である。しかしそのためのデータ獲得や、モデル構築は、基本的に手作業に依存している。そのため広範囲な空間のモデル化には莫大なコストが必要であり、応用分野が限定されたり、対象物を限定せざるを得ないのが現状である。また変化の激しい地域では、一旦構築されたモデルの更新の効率化も問題になる。

この研究では、道路を走行する車両から得られる動画画像を解析することにより、都市景観の3次元情報とテクスチャ情報を効率的に獲得することを目標としている。また獲得された情報を、数値地図を核とするGIS(地理情報システム, Geographic Information System)に投入・組織化することにより、従来の数値地図の表現力と即時性を大幅に高めた3次元・マルチメディア地図へと展開することを計画している。

この報告では、研究の枠組を紹介するとともに、3次元情報獲得手法の詳細と予備的な実験結果を紹介する。

2 都市景観の3次元モデル

都市を構成する物体、すなわち建築物の3次元モデルは、主に3次元グラフィックスの対象として扱われてきた。土木・建築分野においては、グラフィックス技術を利用した景観シミュレーションやウォークスルーシステムのための基礎データとして、主に手作業による入力が行われている。

また自動車の車載装置として定着した感のあるナビゲーションシステムでは、従来からの地図表示による提示だけではなく、ルートに沿った景観の疑似3次元表示や、マルチメディア化が進んでいる。これは地図には「読む」という作業が必要だが、グラフィックスによる鳥瞰的な提示により、一瞥性が向上することが大きく作用している。この例でも、提示のための基礎データである数値地図作成には、膨大な現地調査、手作業による入力を要しており、地図にとっての重要なファクターである即時性を失うことにつながっている。

3次元ビジョン分野では、従来から能動的な観測による3次元情報獲得技術が数多く提案されてきた。最近では、コンピュータの処理速度の向上を背景に、多視点画像や系列画像の冗長性を有効に利用した技術も成熟しつつある[1][2]。しかしその多くは、都市景観のような広い範囲を対象とした場合には、必ずしもうまく適用できるとは言えないのが現状である。

一方、グラフィックスの分野では、高度な幾何演算機能(Geometric Engine)を搭載したレンダリングハードウェアの普及により、アニメーションのリアリティは格段に向上した。しかし、扱える対象物体の数や表面形状の複雑さに限界があることも事実であり、さまざまな提案がなされている[5][6][7]。都市景観のように、対象の存在範囲が非常に広い場合にも、レンダリングすべきオブジェクトの扱い方にそれなりの工夫を要する。

このような背景を考慮し、この本研究では、都市景観の3次元およびテクスチャ情報の獲得に動きステレオの一手法であるEPI(Epipolar Plane Image)を基本とするアルゴリズムを採用し、得られた奥行き情報(Depth Map)と、テクスチャ情報(画素値)を一体化した時空間平面画像として構造化する。この画像は3次元空間の線、面、物体などの認識を行う前のデータであり、物体後との操作は不可能だが、視点の移動に伴う視差を生じさせる提示は可能である。

3 EPI解析による3次元解析

EPI解析は、視点の移動によって生じる(運動)視差の検出を、時空間画像空間に含まれる特殊な平面(エピポーラ平面)画像上での直線検出問題に転換する方法であり、多視点ステレオ、あるいは動きステレオの特殊なケースと見なすことができる[3]。

エピポーラ平面上に描かれる直線の勾配は、対象とする3次元空間の点の奥行きに比例する。通常のステレオでは、離散的な視点におけるステレオペア画像から対応点を求める必要があるが、これに比べると、比較的頑健なステレオ視アルゴリズムであり、「時間軸方向の冗長性を利用した3次元解析手法」といえる[2]。

物体の奥行きを獲得する基本的な方法としては、ステレオマッチング法がある。図1の場合、特徴点 p までの奥行き D は $\Delta U = u_2 - u_1$ とすると式(1)によって求めることができる。

$$D = f \frac{V \cdot \Delta t}{\Delta U} \quad (1)$$

ステレオマッチング法では対応点探索をいかに精度良く行なうかが問題であり、多くの研究が行なわれている。この対応点探索の問題を回避する方法として時空間画像解析による手法が提案されている[3]。時空間画像とは、視点を運動させながら短い時間間隔で撮影した動画画像を時間軸方向に並べて3次元で表現したものである。カメラ運動が等速直線運動のとき、ステレオ立体視におけるエピポーラ線で時空間画像を切断

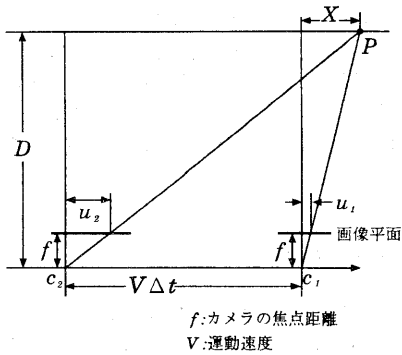


図 1: ステレオマッチング法の原理

したエビポーラ平面画像 (EPI:Epipolar Plane Image) 上に特徴点の軌跡が現れる。

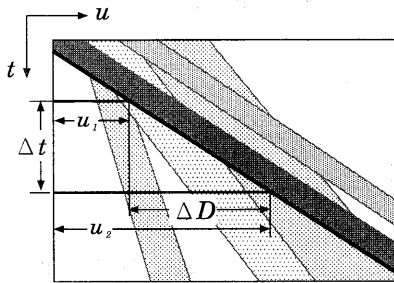


図 2: EPI 上の特徴点の軌跡

特にカメラ光軸がカメラバス(進行方向)に垂直のとき特徴点の軌跡が直線として現れる(図2)。この直線の傾き($a = \Delta U / \Delta t$)を式(1)に代入すると式(2)になり、特徴点までの奥行きが直線の傾きと視点の運動速度のみで求めることができる。

$$D = f \cdot a \cdot V \quad (2)$$

直線検出のみで特徴点までの奥行きが求められるので、エッジ抽出や領域分割などの処理が使え、安定した奥行き計測が可能である。

しかし、カメラ光軸をカメラバスに垂直に設定するため、カメラバスの上・左右の3方向の奥行きを求めるためには3方向の時空間画像が必要となる。そこで、本手法では「拡張エビポーラ平面画像」(EEPI: Extended

EPI)として、カメラ光軸とカメラバスを平行に設定した時空間画像から作るEPIを使用した。カメラ運動が直線運動を行なうとき、EPIにおけるエビポーラ線は静止物体のフロー(画像上での見かけの動き)に沿ったもので、画像平面上では水平な線になる。本来カメラが静止した対象物の世界を移動するとき、画像上に投影された物体のフローはFOE(Focus of Expansion)と呼ばれる一つの特定の画像平面上の点から放射状に湧きだすため(図3)、エビポーラ線は画像平面上のFOEから放射状に伸ばした線になる。

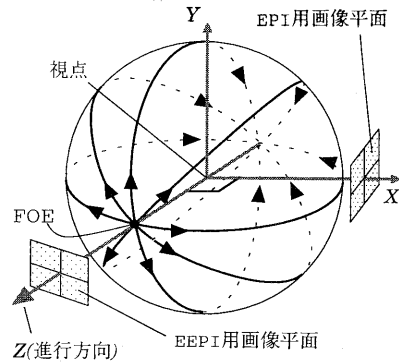


図 3: 直進運動における静止物体の流れ(フロー)

そこで画像平面上のFOEから放射状にスリットを設置して(図4)、時空間画像をスリットで切断したEEPI(図5)(図6-上)を作成する。このEEPIを式(3)を用いてEPI(図6-下)に変換する。

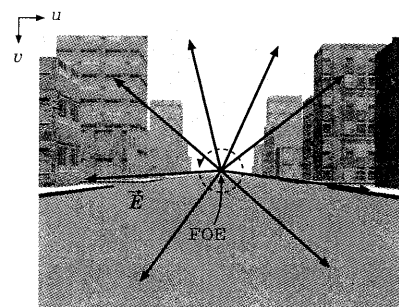


図 4: EEPI 用のスリット

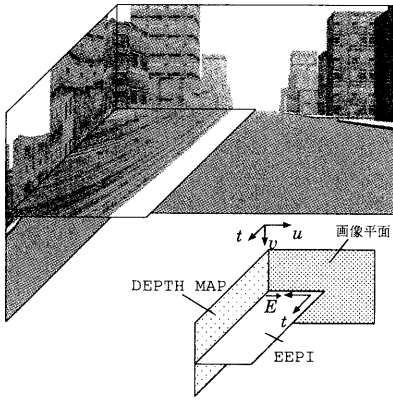


図 5: EEPI と DEPTH MAP

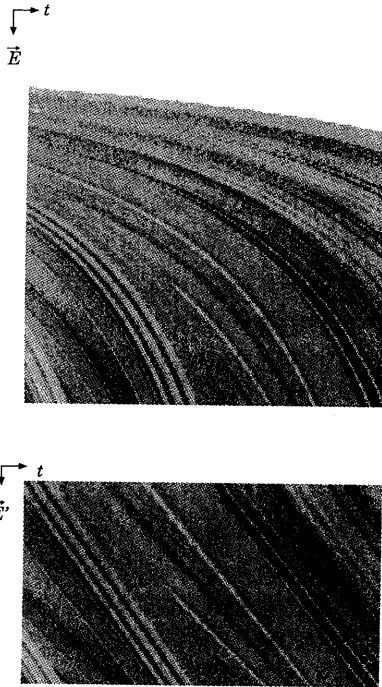


図 6: スリット時空間画像 (上) と変換後の EPI(下)

$$\vec{E}' = f \frac{\vec{E} \cdot \cos \theta + f \cdot \sin \theta}{f \cdot \cos \theta - \vec{E} \cdot \sin \theta} \quad (3)$$

この EEPI 解析では放射状の直線を量子化された画像上で表現するため、奥行き計測精度は従来の EPI 解析より劣るが、1つの時空間画像だけで3方向(左右, 上)の奥行きを求めることが可能になる。また従来の EPI 解析ではカメラパスに平行な面しか計測できないが、EEPI 解析では平行以外の面の計測も可能である。

4 データの構築

計測された奥行きデータから DEPTH MAP を生成する。DEPTH MAP は画像平面上の左右と上の端に3本のスリットを設けて作成したスリット時空間画像(図5)(図7)の1画素ごとに奥行きデータに対応させたものである。画像平面上の全ての静止物体は、無限遠点(FOE)から現れて画像の外に移動していくので、DEPTHMAP 画像用のスリットを必ず通過する。このことから、DEPTH MAP とスリット時空間画像によるデータ構築は3枚のスリット時空間画像のみで画像平面に投影された静止物体を構築することができるため、データ量が大幅に削減することができる。

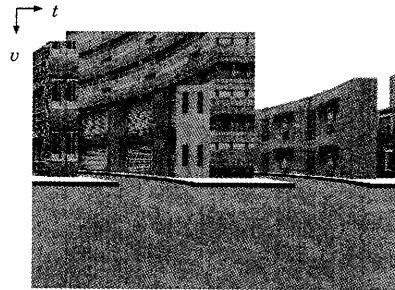


図 7: DEPTH MAP 用スリット時空間画像(カメラパスの左側)

実画像からのデータ獲得と構築が困難である理由の一つに求められた物体の形状を認識することであり、対象が屋外の場合はオクルージョンがおこるため更に複雑になる。また、一般に物体の形状を認識するためには対象物体の全周囲についてのデータを獲得する。そのため、対象物の形状が複雑であるほど困難になる。さらに対象が屋外の場合は、対象物体の大きさや立地状況などから全周囲のデータ獲得が困難である。しかし、本研究では地上を走行する車両から計測できる範囲内に都市空間に存在する対象物体の面のみのデータ

を用いた都市空間の構築を目的としているため、実世界を構成する建物などの完璧なデータを必要としない。また、この DEPTH MAP では 1 画素ごとに奥行きデータを与えられるので、対象物体の形状などを認識する必要がなく、DEPTH MAP から 1 画素ごとに透視変換を行なうことにより仮想都市空間を表現することができる。

この DEPTH MAP では t 軸に平行な 1 ラインが 1 枚の EEPI に対応するため、DEPTH MAP の精度は EEPI 用に設置したスリットの間隔に左右される。間隔を広くすると 1 画素ごとにデータを対応させるためには EEPI 間のデータの補間を行なう必要があり、補間のために EEPI 解析やスリット時空間画像上で“面”を認識しなければならない。EEPI 解析での面の認識には直線検出以外にオクルージョンについての解析 [4] を行なうため処理が複雑になる。そこで実験では画像の幅と同数のスリットを設置した。

これにより EEPI 上で t 軸に平行でかつ \vec{E} 座標が最大 (図 6 では画像の一番下) の 1 ライン上の各画素と、DEPTHMAP の t 軸に平行な 1 ライン上の各画素が対応する。つまり 1 本の直線から求められる奥行きデータが DEPTHMAP 1 画素に対応することになる。1 画素ごとに奥行きデータを与えるので EEPI からの直線検出時に面の認識を行なう必要がなく、処理を容易にすることができる。

この DEPTH MAP を生成することにより、1 本のカメラバスに対するデータを構築することが可能である。しかし、基本的にこのカメラバスは直線運動を想定しているため、1 本の道路分のみである。また奥行きが計測されるのは画像平面に投影される物体のみなので、車両から見えない面などの存在が問題になる。そこで都市空間のデータ構築を行なうためには複数本のカメラバスについて DEPTH MAP を生成し、DEPTH MAP 同士の関係 (交差する位置・角度) を求めることが必要となる。

5 計測結果

実験では車載カメラからの画像を想定して、図 4 を含むシミュレーション画像 (640 × 480pix) 250 枚を作成し、EEPI 解析を行なった。EEPI は左右方向にそれぞれ 480 枚、上方向に 640 枚生成した。

図 8 と図 9 は左方向 480 枚中 1 枚の結果である。図 8 は直線の検出結果で白いラインが検出された直線を表し、図 9 はその直線から求められた奥行きを表示したものである。図 9 中の Z 軸はカメラバス、 X 軸は進行方向左手、四角形は建物、“×” は計測点である。直線

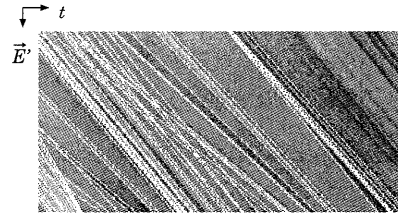


図 8: EEPI からの直線検出結果

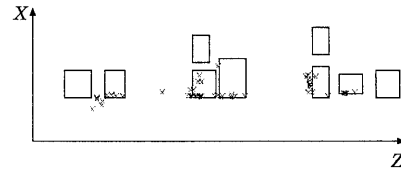


図 9: EEPI からの奥行き計測結果

の誤検出による計測誤差があるが、カメラバスに平行な面でない面についての計測も行なうことができた。

今回使用した直線検出方法は EEPI にエッジ処理を施し、エッジ点について HOUGH 変換を行なったものである。しかし、DEPTH MAP を生成するための検出すべき直線には以下の特徴がある。

- 静止物体は EEPI 上では必ず右下がりの直線になる。
- \vec{E} 座標の最大値の点を含む直線のみが DEPTHMAP に反映される。

このことから単純な HOUGH 変換などでなく、直線の追従処理を行なうことで更に精度良く直線の検出が可能であると考えられる。

以上の方法で獲得、構築された 3 次元情報を投影した結果を図 10, 図 11 に示す。図 10 は右上から斜め左下をふかした時の画像を示す。図 11 は左前方を見ながら前進した時の画像である。

6 奥行きデータの補間

この画像作成に用いた DEPTHMAP には、EEPI 上に特徴点の軌跡として現れた点のみの奥行きしか格納されていない。DEPTHMAP はスリット時空間画像上の画素全てに対して、カメラバスからの奥行き値を与

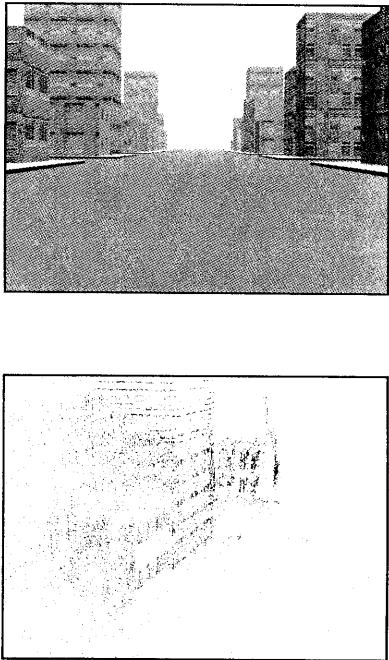


図 10: 再構成画像

える物である。そこで、この計測された箇所の奥行き値を用いて、スリット時空間画像上の全ての画素に対し奥行き値の補間が必要となる。

まず、スリット時空間画像に対して、画像の画素値を用いて領域統合を行った。図 13には、領域分割結果の一例を示した。続いてこの統合された領域データと不完全な DEPTHMAP のデータを用いて補間を行う。補間の手順は次のようにした。

1. 領域分割されたある領域において左右両端の計測点を見つける。
2. それぞれ左右の見つかった計測点において、スリット時空間画像を縦方向にみて領域内で最上と最下の計測点の間で平均値を用いて補間を行う。

この研究の場合、都市空間を対象にしていることから、対象物は地上から垂直に建っているものとみなし、縦方向の相関関係が強いため、縦方向の補間を行っている。

3. 領域内で縦方向に補間した間で、さらに横方向に補間を行う。この際に、縦方向の補間をした左右

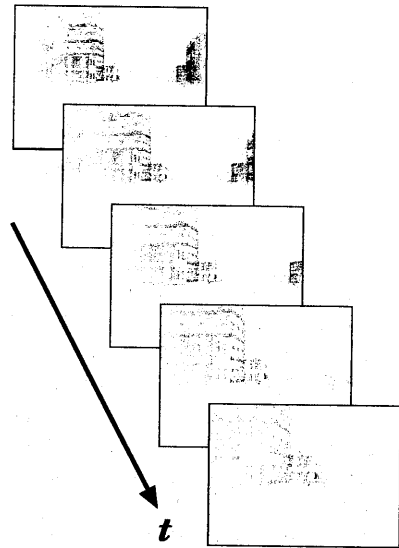


図 11: 再構成画像 2

両端の奥行き値の差が大きい場合には、カメラパスに対して垂直な面とし、差が小さい場合には、カメラパスに対して平行な面とみなした。この条件を元にして、それぞれの場合に対して線形補間を行う。

補間後の DEPTHMAP を用いて再合成した画像を図 14に示す。

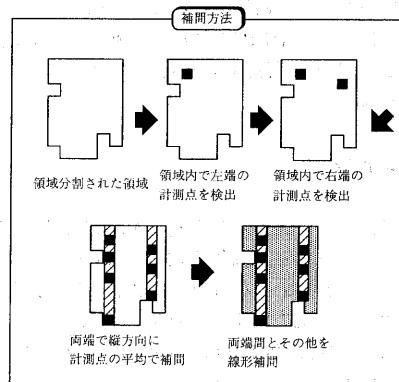


図 12: 補間の過程

7 まとめ

この報告では、仮想化都市空間構築の枠組を紹介し、その基盤技術としての動画像解析手法を提案した。この手法は、EPI解析を基本とし、都市景観の3次元情報とテクスチャ情報を効率よくモデリングすることを目標としている。また、アルゴリズム検証のための基礎的な実験を行った。

今後、以下の各項目に関して検討を進める予定である。

- 実際の走行車両からの実画像に対する実験と精度の向上。
- 既存の情報、たとえば数値地図に含まれる建築物などの地物の幾何学的情報の利用。
- 地上からの観測を補うために、空中（航空）写真やリモートセンシング画像など、異種センサ情報を援用。

この研究の一部は、平成8年度科学研究費補助金（課題番号 08750418）「動画像解析による仮想化都市環境構築に関する研究」による。

参考文献

- [1] C. Toamsi and T. Kanade, “Shape and motion from image streams under orthography: A factor-

ization method”, International Journal of Computer Vision, Vol. 9, No. 2, pp.137-154 (1992).

- [2] 田村秀之, 池田克夫 編, “知能情報メディア”, 第7章「コンピュータビジョン技術を用いた3次元画像メディア」(大田友一), 総研出版, pp. 161-169(1995).
- [3] R. Bolles, H. Baker and D. Marimont, “Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion”, International Journal of Computer Vision, Vol. 1, No. 1, pp.7-55 (1987).
- [4] 安野貴之, 鈴木智, “時空間画像の遮蔽解析を用いた面構造復元”, 情報処理学会論文誌, vol.34, No.10, pp.2174-2183 (1993).
- [5] 北嶋克寛, 遊佐洋子, “リアルタイム景観シミュレータのための形状のグルーピングと多重表現に基づく描画時間の短縮”, 電子情報通信学会論文誌, vol.J77-D-2, No.2, pp311-320 (1994).
- [6] 今村和彦, 安田孝美, 横井茂樹, 鳥脇純一郎, “仮想空間操作に基づく都市景観構築システム”, テレビジョン学会誌, vol.49, No.10, pp1324-1330 (1995).
- [7] 玉田隆史, 中村泰明, “多次元データ構造に基づく3次元仮想都市空間の管理と高速描画”, 電子情報通信学会論文誌, vol.J78-D-2, No.8, pp1205-1213 (1995).

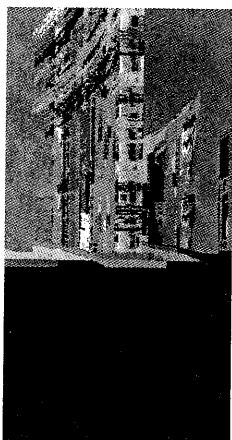


図 13: 時空間画像の領域分割結果

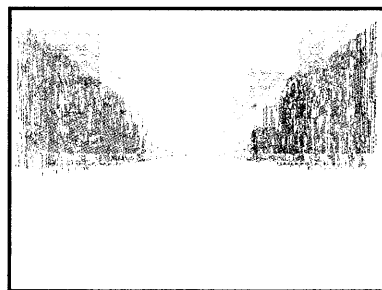


図 14: 補間後データの再構成