

2000 年度 修 士 論 文

仮想空間への実世界情報の投影
Modeling Real-world Activities in Virtual Worlds

村 尾 真 洋
Murao, Michihiro

東京大学大学院新領域創成科学研究科
環境学専攻 社会文化環境コース

目次

第1章	序章	1
1.1	はじめに	2
1.2	研究の目的	3
1.3	ITSにおける位置付け	3
1.4	論文の構成	5
第2章	実世界情報投影の仕組み	6
2.1	はじめに	7
2.2	移動物体の検出	7
2.2.1	背景差分法	7
2.2.2	フレーム間差分法	8
2.3	移動物体の追跡	8
2.3.1	勾配法	8
2.3.2	ブロックマッチング法	9
2.4	投影手法	10
2.5	まとめ	11
第3章	3次元都市モデルの作成	12
3.1	はじめに	13
3.2	既存の技術	13
3.3	幾何学形状モデルの作成	14
3.4	テクスチャマッピング	16
3.4.1	モザイクング	16
3.4.2	広範囲画像の取得	18
3.5	結果	18

目次	2
第4章 プロトタイプシステム	27
4.1 はじめに	27
4.1.1 実験内容	27
4.1.2 実装	27
4.2 車両の検出・追跡	29
4.3 仮想空間への投影	30
4.4 視覚化	32
4.4.1 ガウシアンフィルタ	32
4.5 結果	33
第5章 結論	38
参考文献	40

目 次

1.1	本研究の概念図	3
3.1	ゼンリン住宅地図	14
3.2	3次元都市の幾何学形状モデル	15
3.3	Image mosacing	17
3.4	情報収集車両	19
3.5	機構設計図	20
3.6	取り付けられたビデオカメラ	21
3.7	3台のカメラを用いて取得した画像	22
3.8	垂直方向にモザイクングを行った画像	23
3.9	水平方向にモザイクングを行った画像	24
3.10	幾何学形状モデルとテクスチャマッピングによりモデリングされた3次元都市モデル	25
3.11	図 3.10 にある建物を含む実世界の写真	26
4.1	監視カメラ設置状況	28
4.2	背景画像	31
4.3	トラッキングの結果	31
4.4	車両の追跡結果を仮想空間座標系へ投影した結果	34
4.5	カメラから得られた入力画像	35
4.6	実世界のカメラと同じ位置, 姿勢を再現したビュー	35
4.7	鳥瞰映像	36
4.8	運転席からの眺め	37

第1章 序章

1.1 はじめに

仮想現実 (VR : Virtual Reality) システムの多くは、幾何学形状モデルを用いて、仮想環境や物体を表現し、コンピュータグラフィックス (CG : Computer Graphics) 技術によって描画・表示される。VR の利用分野は広く、シミュレーション、医療応用、福祉応用、3次元モデリング、娯楽、芸術など多岐にわたる。CG 技術の発展に伴い最近では種々のレンダリング技術の進歩によって、実写に近い表示が得られるようになってきている。

また仮想空間におけるモデルについても各種のセンサーやコンピュータビジョンなどの技術を用いて精度の期待できるモデリングが可能になりつつある。仮想空間における動的なオブジェクトの動きに関しても物理法則や動的特性などを考慮し、現実世界で起こり得る動作を仮想空間の中で実現することができるようになった。しかしその運動自体はリアルに表現できるものの、自律的なものの動きは実世界での状態を反映していない場合がほとんどである。例えば、テレビゲームでは主人公のキャラクタはユーザ自身が操り、ドライビングシミュレータにおける周囲の車の走行の仕方はランダムであったり、またはある一定の決められた規則に従って動作していることが多い。そこで本研究では実世界における活動を仮想空間に投影することを目的としている。これはいま現在進行している実世界での活動を実時間のうちに3次元の仮想空間へ挿入することを意味する。図 1.1 はこの研究の概念を示す。

本研究における仮想空間とはCG技術を利用して電子的に再現した3次元空間を意味する。また、実世界情報を採取する方法としては各種センサを用いることが考えられるが、本研究においては手段としてカメラに代表される視覚センサによって実世界情報を取り込み、仮想3次元空間の中で活動を再現し、利用者に視覚的に提示すること考える。

実世界の情報を仮想空間へ投影することの利点は、任意の視点から対象となる空間を眺められることである。これにより実世界における対象オブジェクトと、その周囲の環境との空間的な関係を容易に把握できるようになる。

このような視覚化が有効な例としては、スポーツシーンでの多地点からの観戦、監視システム、ITS (Intelligent Transportation Systems : 高度道路交通システム) などへの適用などが考えられる。スポーツシーンでは視点を変えて表示することにより、プレイヤーの迫力を味わうことができるようになる。またデータ分析にも有用である。監視システムでは工場、倉庫、プラントなどの管理区域、危険区域などで役に立つものと思われる。本研究ではITSでの利用を考慮し、交通状況の視覚化によるドライバーの安全運転支援を目的とするシステムを提案する。

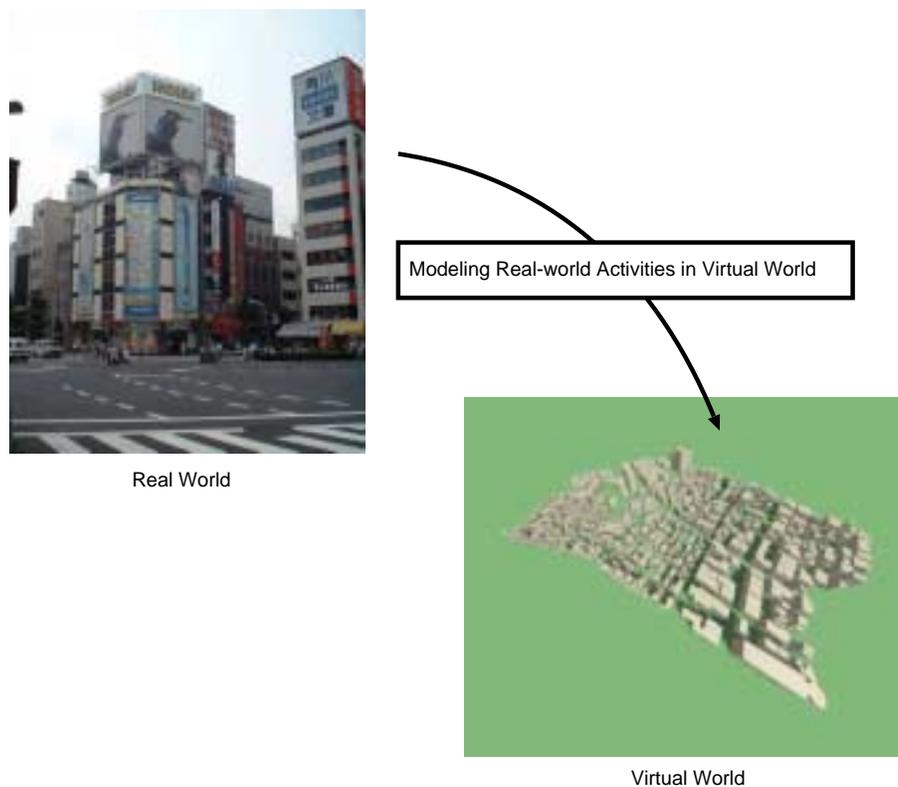


図 1.1: 本研究の概念図

1.2 研究の目的

実世界情報を仮想空間に投影することを目的としている。本研究では特に実世界情報として視覚情報を用いる。視覚センサであるカメラを通して2次元画像を採取しコンピュータビジョンの技術により動物体を検出・追跡する。その結果を仮想空間へ投影し、結果をCGにより表示する。実世界情報のうち、建物など物体に関してはあらかじめ幾何学形状モデルを持っておくことを前提とする。仮想空間では

またようなシステムの実装を行うための技術的解法について考える。

1.3 ITSにおける位置付け

ITSは道路交通に関する総合的な情報通信システムであり、人と道路と車両とを情報で結び道路問題を解決する。その目的は、交通の安全性の向上、円滑化、および環境改善による快適性の向上である。ITSを構成する開発分野は以下の9つに大別されている。

1. ナビゲーションの高度化

2. 自動料金收受システム
3. 安全運転の支援
4. 交通管理の最適化
5. 道路管理の効率化
6. 公共交通の支援
7. 商用車の効率化
8. 歩行者などの支援
9. 緊急車両の運行支援

現在,日本ではこれらのシステムの一部は実現されつつある.例えば,1 に関しては VICS (Vehicle Information and Communication System) というシステムが開始されており,2000 年 12 月現在 250 万台のユニットが出荷されている.VICS によりドライバは交通関連情報をリアルタイムで受信することができる.このおかげで渋滞情報や最短経路等がナビゲーションシステムに表示され,移動時間の短縮により快適な移動が可能となる.また,2 においては 2000 年 4 月から首都圏の主要な料金所で道路自動料金收受システム (ETC: Electronic Toll Collection System) の試行運用が開始されており,料金所の渋滞解消や,利便性の向上が見込まれている.2001 年度末には全国約 800 カ所の料金所に拡大され,2002 年度までに全国の主要な料金所約 900 カ所に導入される予定である.

日本の車両保有台数は年々増加傾向にあり,1998 年末の保有台数は約 7081 万台となっている.自動車はいつでもどこへでも自由に移動できるという利便性や快適性において他の交通手段にない特徴をもっており,これによって現在ような自動車社会が生まれ出された.しかし他方では,自動車交通は交通事故による多数の死傷者を発生する原因となった.日本国内においては 1970 年の 16765 人を最高として 1980 年ごろまでにほぼ半減し,その後の横ばい・増加を経て最近では漸減の傾向にある.しかし,今なお事故件数は年間 80 万件,負傷者数は 100 万人を超え安全性の向上が望まれている.

このような状況の中,ITS の開発分野においては 3 にあるドライバに対する安全運転支援は重要視されており,効果的なシステムの開発が望まれている.交通事故の原因を過去の事故発生内容から分析した結果として,ドライバの発見の遅れと判断の誤りが全体事故の中で 60 このことから事故を低減するためにはドライバに必要な情報を適宜提供すること,ドライバへの警告情報の提供・危険回避のための緊急操作支援を行うことなどが考えられる.ところで本研究の主題である仮想空間への実世界情報の投影の利点は,状況を視

覚化しユーザに任意視点からのビューを提供できることにある．これによりたとえば実世界情報として道路状況を採取し，仮想空間の中でそれを再現することによりドライバへ任意の視点から見た交通状況を提供するという利用が考えられる．交差点付近などでは特に事故件数が多いとの報告がなされており，交差点での右折事故や出会い頭の衝突事故などへも有用な一助になると予想される．

[1] は見通し外交差点の出会い頭事故防止を目標に，電波の回折による車車間通信を用いてサービスエリア内のすべての車に対し，互いの車の位置，速度や衝突が回避できる距離などの情報を知らせることにより事故を未然に回避しようとするものである．これに対し本研究では交差点などに設置されているインフラとしての固定カメラの映像を利用することで安全支援に役立てることを提案する．安全運転の支援では，事故などを未然に防ぐために，道路や周辺車両の状況など走行環境状況をドライバーに視覚的に提供することが有効である．そこで，インフラとして設置されている視覚センサーから得られた映像をもとに，時々刻々と変化する道路状況を仮想空間の中に再現し，実時間でドライバーに提示するシステムが考えられる．このようなシステムは，ドライバーの望む任意視点からのビューを提示することができるし，またCGによる視覚的效果を付加することも可能である．例えば，交差点において，ドライバーの死角となる建物などの陰を走る車を表示することができ，これは出会い頭の衝突防止支援となる．

固定カメラによって得られる映像をもとにしてドライバーに任意視点からのビューを提示するには，画像に対して射影変換を適用して入力画像とは異なる視点から見た場合の画像を生成する写実的な方法 [2] と，表示するモデルすべての幾何情報を持っておき，CGによってビューを表示する幾何的な方法の二通りが考えられる．しかし前者の場合，一視点から得られる画像を変換して完全に任意の視点から見たような画像を生成することは明らかに不可能である．ここでは，後者の方法によりビューを生成することを考え，以降の章においてこのようなシステムを実現するための方法を考えていく．

1.4 論文の構成

第 2 章では実世界情報を仮想空間へ投影する仕組みについて説明する．

第 3 章では 3 次元都市モデルをどのように作成するかについて論じる．

第 4 章において実世界情報を投影するようなプロトタイプシステムについて説明を行う．

第 5 章では第 2 章から第 4 章を要約する．

第2章 実世界情報投影の仕組み

2.1 はじめに

本章では実世界情報を仮想空間へ投影するための仕組みについて論じる。

前章で述べたとおり，研究ではカメラによって実世界情報を捕らえ，そこから得られる情報を分析し，視覚化することを考える。

投影の流れは次のとおりである．まず実世界情報をカメラによって得る．ここから動物体を検出する．次にその動物体の活動を仮想空間へ投影する．最後に仮想空間を表示することで，結果を得ることができる。

ここではそれぞれに必要な技術に関して技術的な説明を行う．以下，2.2 節にて移動する物体をどのようにして検出するかを説明する．次に 2.3 節において，フレーム間で移動物体を同一のものとして追跡する手法を説明する。

なお結果として CG によって仮想空間を表示するには，実世界に存在する物体の幾何学形状モデルを持っておく必要がある．幾何学形状モデルの作成に関しては第 3 章において後述する。

2.2 移動物体の検出

車両領域を検出する代表的な方法としては，あらかじめ作成しておいた背景画像と着目する画像との間において座標で対応する各画素について差分をとることで車両領域のみを検出する背景差分法や，連続するフレーム間で同様の差分をとることで変化のあった部分を検出するフレーム間差分法が挙げられる。

2.2.1 背景差分法

背景差分法を用いた方法は，撮影系の移動物体が存在しない画像（背景画像）を用意し，それと入力画像との差分をとることで移動物体の存在する領域を検出する手法である．判別式 (2.1) で， $B(t)$, $i(t)$, $th(t)$ はそれぞれ時刻 t における背景カラー値，入力カラー値，閾値である．このような比較を各ピクセルに対して適用することで移動物体の存在領域を判断する。

$$|B(t) - i(t)| < th(t) \quad (2.1)$$

閾値は動的に更新されたり，ピクセルごとに変化させてより精度よく移動物体を検出する研究もなされている．背景差分法を用いる場合，正確な背景画像を獲得することができれば，ほぼ的確に移動物体の存在する領域を検出することができる．しかし，実世界環境下においては，照度の変化などにより常に背景が変動し続けるため，それに追従することが必須となる。

2.2.2 フレーム間差分法

フレーム間差分法 [8] は、連続するフレーム間で対応する各画素について差分を求めるものである。時刻 t における着目画素の画素値を $i(t)$ 、1 フレーム前の同位置の画素値を $i(t-1)$ とし、閾値を $th(t)$ とすれば、式 (2.2) で表される。

$$|i(t) - i(t-1)| > th(t) \quad (2.2)$$

このフレーム間差分法を移動物体検出に用いる際には、フレーム間差分の絶対値が大きい部分を以下のいずれかであるとして検出する。この中のどの状態であるかはフレーム間差分だけでは知ることができない。

- 移動物体が着目画素に侵入した
- 移動物体が着目画素から出た
- 照度が変化した

フレーム間差分法は計算コストが低く、比較的高速に動く物体などを検出する際には有効である。

2.3 移動物体の追跡

移動物体の領域検出により、各フレーム内での移動物体の存在領域がわかるが、フレーム間においてその移動物体が同一であることを知るために移動物体のトラッキング (追跡) が必要となる。移動物体のトラッキングができれば、その移動物体の動きの軌跡がわかるばかりか、算出により速度も得られ、さらに次の瞬間にその物体がどの位置に来るのかということを予測することも可能である。移動物体のトラッキングの代表的な手法としては、オプティカルフロー [9] が挙げられる。

オプティカルフローとは画像の輝度パターンに基づいて計算されるイメージフローの推定値である。つまり、オプティカルフローは画像上の各点の速度ベクトルとなる。オプティカルフローの推定法は大きく勾配法とブロックマッチング法に分類される。

2.3.1 勾配法

勾配法は見掛けの動きが小さい場合に用いられる方法である。画像上の点 (x, y) の時刻 t における輝度を $I(x, y, t)$ とする。物体上の同一点の観察される輝度が時間的に不変で

あると仮定すると，画像上の各点 (x, y) で近似的に以下の拘束式が成り立つことが知られている．

$$I_x u + I_y v + I_t = 0 \quad (2.3)$$

ただし， $(u, v) = \left(\frac{dx}{dt}, \frac{dy}{dt}\right)$ は各点での速度ベクトルである．ここで， I_x, I_y, I_t は動画像から差分近似により計算でき，速度ベクトルの 2 つの自由度のうち一つが減ったことになる．残りの自由度を決定し，速度ベクトルを求めるためには，対象の動きに対する仮定が必要である．例えば，2 点が同じ動きをしていると仮定すると，連立方程式を解くことによって (u, v) が定まり，多数の点と同じ動きをしているならば最小二乗法で (u, v) を決定することができる．

勾配法を用いた手法は，対応点検索をおこなう必要がない利点があるが，輝度値が急激に変化するところではフローの誤差が激しいこと，また雑音に弱いと言う欠点がある．

2.3.2 ブロックマッチング法

ブロックマッチング法は画像を一定の大きさの小領域 (ブロック) に分割し，それぞれのブロックの動きベクトルを求める．これは各ブロックが前のフレームのどこに対応するかを探し，対応するブロックの位置の差を動きベクトルとするものである．

ブロックマッチング法では以下の計算を行う．いま第 t フレームの $N \times N$ の大きさのブロックの動きベクトルを求めるものとする．ブロックの左上端の画素の位置を (i_0, j_0) とする．このブロックの画像と直前のフレームの対応する位置から (p, q) だけずらした位置のブロックの画像との差を評価する．マッチングの評価には以下の式に示すような，画素値の差の絶対値の和 (SAD : Sum of Absolute Difference) などが用いられる．

$$SAD(p, q) = \sum_{j=0}^{N-1} \sum_{i=0}^{N-1} |f_t(i_0 + i, j_0 + j) - f_{t-1}(i_0 + i + p, j_0 + j + q)| \quad (2.4)$$

対応候補のブロックの位置 (p, q) を変えて $SAD(p, q)$ を計算し，最小となる (p, q) を動きベクトル (p_m, q_m) とする探索範囲 R_s は， $-l \leq p, q \leq h$ の正方形であることが多い (l, h は正数)．以上を式で表せば，

$$(p_m, q_m) = \arg \min_{(p, q) \in R_s} SAM(p, q) \quad (2.5)$$

となる．なお， $\arg \min_{(p, q) \in R_s} SAM(p, q)$ は $SAM(p, q)$ を最小とする (p, q) を意味する．

ブロックマッチングによって求めなければならない動きベクトルの数はブロックサイズ N が大きいほど少なくなるので，なるべく N を大きくしたいが， N が大きくなるとブロックの中の動きが一様でなくなるため，予測誤差が大きくなる．通常ブロックの大きさ

は 8×8 から 16×16 程度である．また探索範囲は 32×32 程度である．ブロックマッチング法を用いた手法は，輝度値の急激に変化する所でもフローの誤差が少なく雑音にも強いが，探索範囲によっては計算時間が膨大となる．また拡大・縮小，回転運動に弱いという欠点がある．

2.4 投影手法

実世界情報を仮想空間へ投影するという事は，ここでは 2 次元画像中の任意物体を 3 次元仮想空間に投影することであると考えられる．2 次元画像は 3 次元の世界を画像へ投影したものであるといえる．しかし 1 次元分の情報が不足するため，1 つのカメラから得られた画像のみからでは 3 次元物体の形状を一意に決定することができない [10]．このように完全に決定できないことを不定性が残るといふ．そこで一般的には複数のカメラを用いて画像を得る．複数のカメラから得られた画像からは，エピポラ幾何によりカメラ間の位置と姿勢の情報が得られる．さらに対象物の形状に関する情報が得られる．複数のカメラより得られた画像から形状を復元することをステレオと呼ぶ．2 つのカメラの内部パラメータならびに外部パラメータが既知であるとき，対象物の 3 次元形状を不定性なく復元することができる．射影カメラを考えた場合，3 次元中の点 X は 2 次元画像上の点 m に次のように投影される．

$$\lambda \tilde{m} = P \tilde{X} \quad (2.6)$$

ここで， $\tilde{m} = [m_1, m_2, m_3]^\top$ は $m = [u, v]^\top$ に対して $[m_1, m_2, m_3]^\top \sim [u, v, 1]^\top$ なる同値関係が成り立つ同次座標であり， $\tilde{X} = [X_1, X_2, X_3, X_4]^\top$ は $X = [X, Y, Z]^\top$ に対して $[X_1, X_2, X_3, X_4]^\top \sim [X, Y, Z, 1]^\top$ なる同値関係が成り立つ同次座標であるとする．また， P は 3×4 の射影カメラ行列である．式 2.6 を開いて X, Y, Z に関してまとめると次のようになる．

$$\begin{pmatrix} p_{31}u - p_{11} & p_{32}u - p_{12} & p_{33}u - p_{13} \\ p_{31}v - p_{21} & p_{32}v - p_{22} & p_{33}v - p_{23} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} p_{14} - p_{34}u \\ p_{24} - p_{34}v \end{pmatrix} \quad (2.7)$$

カメラの内部パラメータおよび外部パラメータがともに既知であるとするとき，式 2.7 の中の p_{11}, \dots, p_{34} は既知である．また，画像より投影像の座標 u, v が得られるので式 2.7 は X, Y, Z を未知数とする連立一次方程式である．一つの視点の画像からは式 2.7 のように拘束式が 2 つ得られる．したがって，2 つの視点の画像からは次のように 4 つの拘束式が得られる．

$$\begin{pmatrix} p_{31}u - p_{11} & p_{32}u - p_{12} & p_{33}u - p_{13} \\ p_{31}v - p_{21} & p_{32}v - p_{22} & p_{33}v - p_{23} \\ p'_{31}u' - p'_{11} & p'_{32}u' - p'_{12} & p'_{33}u' - p'_{13} \\ p'_{31}v' - p'_{21} & p'_{32}v' - p'_{22} & p'_{33}v' - p'_{23} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} p_{14} - p_{34}u \\ p_{24} - p_{34}v \\ p'_{14} - p'_{34}u' \\ p'_{24} - p'_{34}v' \end{pmatrix} \quad (2.8)$$

この式には未知数 3 に対して拘束が 4 つあるので，これを

$$MX = b \quad (2.9)$$

とおくと， X の最小二乗解は M の一般化逆行列 M^+ を使って次のように求まる．

$$X = M^+b \quad (2.10)$$

3 つ以上の画像が得られている場合にも， M 行列と b ベクトルにこれらを付け加えていくことにより，全く同様に X の最小二乗解が求まる．画像上の全ての特徴点に対してこのように X を求めれば対象物の 3 次元形状を不定性なく復元することができる．

なお，カメラ行列 P を求めるには P に定数倍の不定性があることにより最後の項 p_{34} を $p_{34} = 1$ と仮定して残りの 11 個の要素を決めることができる． m と X が既知であれば式 2.6 より P に関する拘束式が次のように 2 つ得られる．

$$\begin{bmatrix} X & Y & Z & 1 & 0 & 0 & 0 & 0 & -uX & -uY & -uZ \\ 0 & 0 & 0 & 0 & X & Y & Z & 1 & -vX & -vY & -vZ \end{bmatrix} \begin{bmatrix} p_{11} \\ p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{22} \\ p_{23} \\ p_{24} \\ p_{31} \\ p_{32} \\ p_{33} \end{bmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} \quad (2.11)$$

P の自由は 11 であるから，このような点が 6 点あれば十分な拘束が得られてカメラ行列 P を決定することができる．

3 次元物体を撮影した 2 次元画像に基づいて，任意視点からの画像を生成するためには 3 次元形状を復元し，それを出力画像上に投影することになる．

2.5 まとめ

本章では仮想空間への実世界情報の投影に必要な技術について概説した．これらの技術を利用することで実際プロトタイプシステムを作成した．プロトタイプシステムに関しては第 4 章において詳細を述べる．

第3章 3次元都市モデルの作成

3.1 はじめに

本章では仮想空間となる 3 次元都市モデルの構築について述べる。

本研究における仮想空間とは現実の都市空間をもとに CG 技術を利用して電子的に再現した 3 次元都市空間を意味する。この仮想都市の用途として、従来から公共や企業における地下の整備管理や都市計画における景観シミュレーション、またマーケティングや教育などの広範囲な分野での応用が実現されている。これに伴い 3 次元デジタル地図などの空間情報の重要性が高まっているといえる。

よりリアリティを持った都市空間モデルを構築するためには建物表面のテクスチャが不可欠である。そこで、本研究においてはモデル作成の方法として、幾何学形状モデルにテクスチャマッピングを行う手法を採る。以下の節では既存のモデリング技術、デジタル数値地図をもとにした幾何学形状モデルの作成、実世界都市空間におけるテクスチャの取得に関して述べる。

3.2 既存の技術

VR をはじめとする物体の 3 次元表示技術の発展に伴い、都市空間のモデル化もよりいっそう重要になりつつある。3 次元都市モデル作成に関する研究は近年盛んに行われている [16]。3 次元形状モデルを得る手段は様々であるが、代表的な例としては光学センサを用いて航空写真によるステレオ視を行ったり [15] [17]、レーザースキャナを利用して 3 次元情報を得ながら同時にテクスチャも取得し、建物のモデルを構築するといった研究が見られる [18]。

こういった研究により 3 次元デジタル地図は身近なものになりつつあるが、一方で幾何学形状モデル作成の自動化はやはり困難であり作業の効率化と大衆化が求められているのも確かである。そこで、モデリングの難点を解消するため、実写から CG 映像を作成する技術が盛んに研究されている。画像列から新たな画像を直接生成する技術にはモーフィング、QuickTime VR [3]、光線空間法などがあり、Image-based rendering と呼ばれている。Image-based rendering では、複数の撮影条件で撮影された画像から、他の撮影条件で得られるであろう画像を補間によって作り出す。しかし画像ベースで表現すると視点の移動に関して違和感が生じる。また大量の画像データを必要とする。画像の補間によって画像を生成するため、任意視点の移動が可能であるといっても完全ではない、などの問題がある。

幾何学形状モデルを持つか、画像をベースに表示を行うかという選択はその利用方法により適切に選択されるべき技術である。

3.3 幾何学形状モデルの作成

都市部でほぼ整備が完了しつつあるデジタル化された 2 次元の地形図を用いて 3 次元建物の幾何学形状モデルを生成する．この生成方法では，住宅地図に含まれる建物図形と階数から立体図形を自動生成するため，詳細さにはかける問題がある．しかし，用によっては単純な形の図形による近似で十分であり，本研究においてもこの方法を採用した．

建物図形に高さを階数 $\times 4$ メートルの壁面と天井面を与えることで，仮想的な建物オブジェクトを構築した．この幾何学形状作成のための 2 次元地図にはゼンリン住宅地図 ZMAP-TOWNII (1/2500, shape file, 東京都千代田区) を利用した．図 3.1 に 2 次元の住宅地図を示す．また，壁面と天井面を与えた 3 次元幾何学形状モデルを図 3.2 に示す．

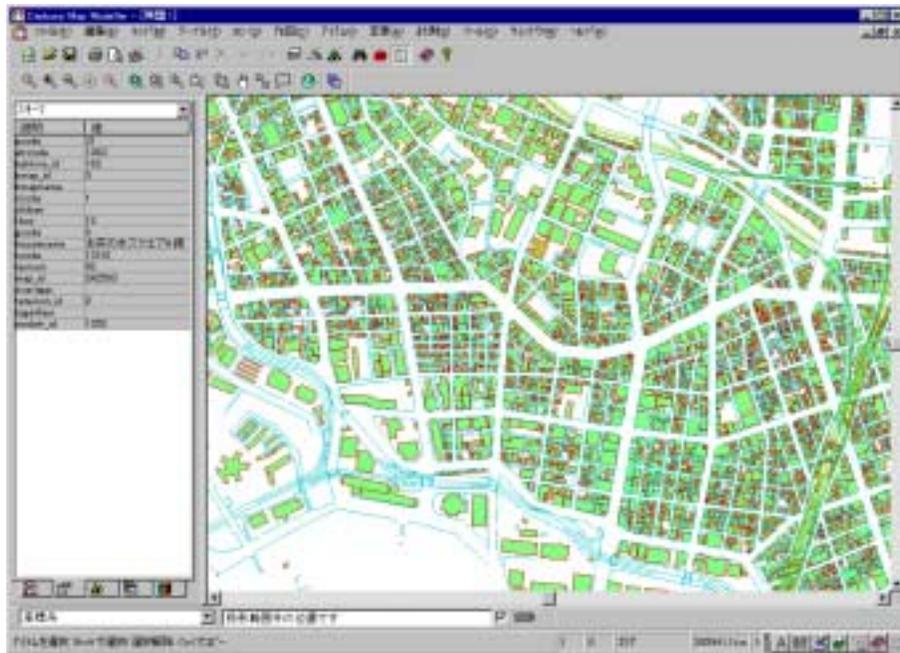


図 3.1: ゼンリン住宅地図



図 3.2: 3次元都市の幾何学形状モデル

3.4 テクスチャマッピング

3.4.1 モザイクング

複数の画像をつなぎ合わせ、視野が広く、高解像度のパノラマ画像を生成する手法をイメージモザイクング (Image mosaicing) という。イメージモザイクングの例を図 3.3 に示す。

この技術は航空写真やリモートセンシング画像の合成などに利用されてきたが、最近では仮想現実感の生成 [3] や監視システム [4]、映像ブラウジングインタフェース [5] といった応用も提案されている。

一般的にパノラマ画像の生成は、平面射影変換により一方の画像を他方の画像と重ねるという手法を採る。これは 2 台のカメラで同一の平面物体を撮影したとき、2 枚の画像間は射影変換で結ばれるという原理に基づいている。また平面物体に限らず、カメラから十分に離れているシーンに関しても平面であると近似することにより、同じ手法を適用することでパノラマ画像を生成できる。画像上の点 $m_1 = [x_1, y_1, 1]^T$ 、もう一方の画像上の点 $m_2 = [x_2, y_2, 1]^T$ 、2 枚の画像を重ね合わせる平面射影変換行列 H の関係は次のように表される。

$$\lambda m_2 = H m_1$$

$$\lambda \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} h_0 & h_1 & h_2 \\ h_3 & h_4 & h_5 \\ h_6 & h_7 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} \quad (3.1)$$

$$\begin{cases} x_2 = \frac{h_0 x_1 + h_1 y_1 + h_2}{h_6 x_1 + h_7 y_1 + 1} \\ y_2 = \frac{h_3 x_1 + h_4 y_1 + h_5}{h_6 x_1 + h_7 y_1 + 1} \end{cases} \quad (3.2)$$

この平面射影変換行列 H の算出がモザイクング技術のキーポイントとなる。Szeliski ら [7] は輝度値の誤差を評価関数として Levenberg-Marquardt 法と呼ばれる非線型反復最小化法により平面射影変換を推定している。Levenberg-Marquardt 法を用いた平面射影変換行列の推定は、評価関数として原画像の画素値 I と変換後の I' との誤差の二乗和 E を用いる。 E を式 3.3 のように定義する。

$$E = \sum_i [I'(x'_i, y'_i) - I(x_i, y_i)]^2 = \sum_i e_i^2 \quad (3.3)$$

そしてこの評価関数 E を最小化するように非線型反復最小法の手法により最適化を行う。この手法の問題点は計算量が多いことと局所解に陥りやすいということである。また、千葉ら [6] は画像上の対応する特徴点を利用して変換行列の算出を行っている。この手法は



(a)



(b)



(c)

☒ 3.3: Image mosacing

まず正規化相互相関を用いて画像の重なり部を抽出する．その後オプティカルフロー推定に基づき画像の特徴点を対応付ける．最後に最小二乗法により特徴点の対応から平面射影変換行列を算出する．このような画像特徴に基づくモザイクング手法では，テクスチャが少ない画像では特徴点の対応が見つけれなかったり，誤対応が生じることが問題となる．

3.4.2 広範囲画像の取得

テクスチャとなる画像の取得はデジタルカメラなどを用いて建物を写すのが一般的である．しかし，建物一つ一つを撮影するには手間と時間が必要となる．そこで，車の上部に複数台のデジタルビデオカメラを設置し，広範囲の映像を取得する手法を採った．この情報取得のための車両の概観を図 3.4 に示す．

一台のカメラで建物全体を一枚の画像におさめるように撮影するには，被写体となる建物がカメラから離れており，かつその建物が高層でない場合に限られてしまう．ここでは 3 台のビデオカメラを角度を変えて設置することにより，建物の上部，中部，下部を撮影した．この機構の設計図を図 3.5 に示す．またこの機構を設置し，天井部にビデオカメラを取り付けた図を図 3.6 に示す．

車両は道路を走りながら，進行方向に向かって右側にある建物のテクスチャを取得する．この機構では 3 台のカメラが固定された状態で映像が取得できるため，上部と中部と下部を写す映像の間にずれが生じない．

3 つに別れた映像から建物のテクスチャを作成する．これには，取得したいテクスチャを持つ建物の面が道路に向かって垂直に立っており，横にならぶ建物の面はすべて同一の面にあると仮定している．実際，都市にはこのようなオフィス街が多く，この仮定はそう厳しくはないものである．

3 つのビデオカメラから取得した画像を図 3.7(a) に示す．3 つの画像それぞれに対して射影変換をかけることにより，面を起こす (図 3.7(b)) ．これら 3 つの画像は垂直方向に平行移動させることにより画像同士が重なり，これにより図 3.8 のような画像が得られる．次にこのような画像を水平方向に張り合わせるにより広域なテクスチャが得られる．広域なテクスチャの画像を 3.9 に示す．このように広域なテクスチャを自動的に取得し，テクスチャマッピングを行うことにより，広い 3 次元都市モデルを比較的簡単に作成することができる．

3.5 結果

以上の方法により，幾何学形状モデルを生成，建物の壁面に取得したテクスチャを張り付けることで 3 次元都市モデルを作成した．



(a)



(b)

図 3.4: 情報収集車両

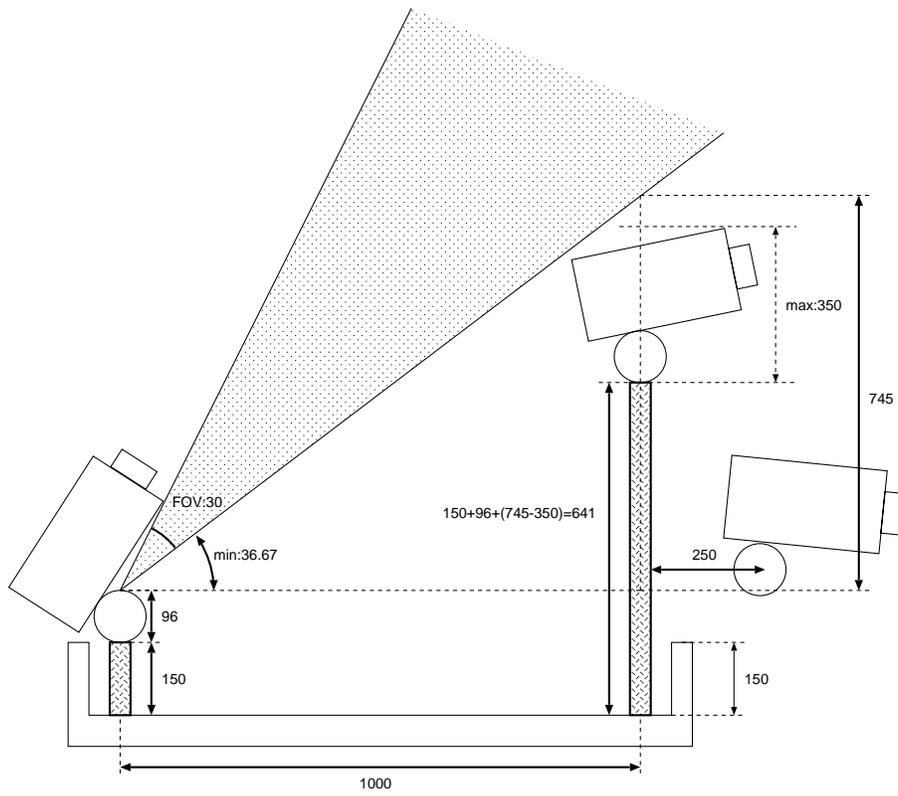


図 3.5: 機構設計図



図 3.6: 取り付けられたビデオカメラ

図 3.10 にモデリングの例を示す．また比較のため図 3.10 にある建物を含む実世界の画像を 3.11 に示す．



(a) 射影変換前

(b) 射影変換後

図 3.7: 3 台のカメラを用いて取得した画像



図 3.8: 垂直方向にモザイクを行った画像



図 3.9: 水平方向にモザイクを行った画像

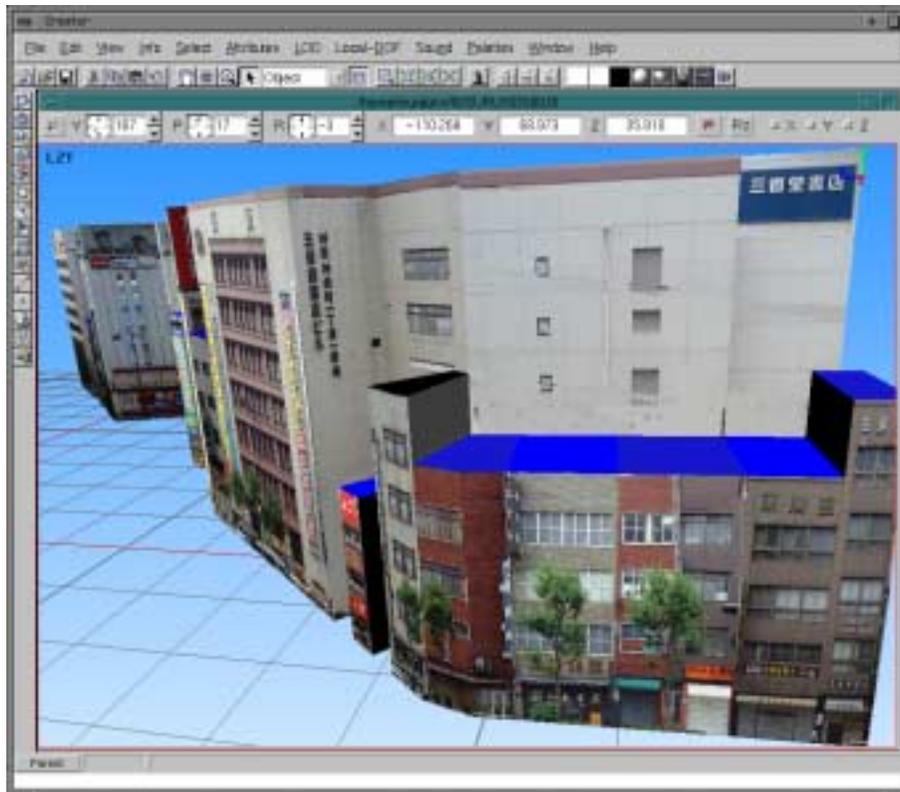


図 3.10: 幾何学形状モデルとテクスチャマッピングによりモデリングされた 3 次元都市モデル



図 3.11: 図 3.10 にある建物を含む実世界の写真

第4章 プロトタイプシステム

4.1 はじめに

ここでは、実際に交差点付近に設置されたビデオカメラから得られる映像を利用し、実世界の交通状況の視覚化を行うプロトタイプシステムについて実装、実行結果を交え説明する。交通状況の視覚化によってドライバは他車の位置や速度を直観的に理解することができる。これは互いの危険を察知することにつながり、安全運転の支援となることが期待される。

4.1.1 実験内容

東京都千代田区駿河台下交差点に面するビルの屋上に設置されたビデオカメラは交差点を監視している、この交差点の特徴は、複雑で交通量が多く、日本で有数の事故数を数える。駿河台下交差点脇の Victoria ゴルフ館屋上 (4.1(a)) に設置してあり、1999年11月から無停止で作動している。設置の目的としては、車両の動きがもっとも複雑になる交差点に着目し、その中で交通事故やニアミスなどの交通イベントの自動検出を行なう技術を開発するためと、大規模な交通画像データベースを構築することである。本実験ではこのカメラから得られる映像をもとに、交差点を走る車両を仮想空間の中で走らせることを目的としている。

4.1.2 実装

プロトタイプシステムの実装には以下のようなステップを適用する。

1. 入力映像からの車両領域の検出・トラッキング
2. 座標系の変換
3. CG によるビューの表示

以下の節においてそれぞれのステップに関して詳細に述べる。



(a) カメラが設置してあるビル (正面)



(b) 監視カメラ

図 4.1: 監視カメラ設置状況

4.2 車両の検出・追跡

動画像における移動物体の検出・追跡 (トラッキング) に関しては第 2 章においてその代表的な手法を示した。本実験では上條ら [14] のアルゴリズムを適用した。一般に、ブロックマッチングの手法では、カメラと車の遠近による車の大きさの変化や、車の見え方の変化によりマッチングが取れないという問題が生じる。[14] では背景画像を適宜更新しながら背景差分法により車両領域を検出し、さらにブロックマッチング法による推定を利用して頑強なトラッキングを実現している。

ここで、上條らのアルゴリズムを説明する。まず解像度 640×480 ピクセルの入力画像を 8×6 ピクセルごとのブロックに分割する。そして 80×60 ブロックごとにどの車両の ID に属するかを決定する。以下、 80×60 ブロックの書くブロックがいずれの車両 ID に属するかを記述するためのマップを Object-Map と呼ぶ。入力画像に対し、以下のステップからなるアルゴリズムを適用する。ここで、いずれかの移動物体 ID に属するブロックを Object-Block と呼ぶ。

1. 初期化

交差点への出入口にスリットを設ける。

2. 移動物体 ID の作成

それぞれのスリットにおいて、現画像と背景画像との画措置の差分が閾値より大きいブロックを検出する。検出されたブロック同士で隣接するブロックをクラスタリングして、新たな移動物体 ID を割り当てる。新たな移動物体がすでに存在する移動物体と重なっているときは、すでに存在する物体の ID に融合させる。

3. 移動物体の動きベクトル推定

移動物体に属する複数のブロックそれぞれにブロックマッチング法を適用し、動きベクトルを求める。時刻 t の画像フレームにおいて $(x(t), y(t))$ にあるブロックと、時刻 $t + 1$ の画像フレームにおいて $(x(t + 1), y(t + 1)) = (x(t) + v(t), y(t) + u(t))$ にあるブロックとの類似度は下式で計算される。

$$D = \sum_{0 \leq i \leq 8, 0 \leq j \leq 8} |I(x(t + 1) + i, y(t + 1) + j) - I(x(t) + i, y(t) + j)| \quad (4.1)$$

ここで I はピクセル画素値を表す。各ブロックについて求められた動きベクトルのうち、最多頻度のものをその移動物体の動きベクトルとして推定する。

4. 移動物体の更新

ステップ 3 で推定した動きベクトルをもとに，現画像上の移動物体を次画像上で移動させる．移動した先では移動物体の画像上での大きさや形状が変化するため背景画像との差分が閾値以下であるブロックから移動物体 ID が削除され，近傍ブロックのうち背景画像との差分が閾値を越えるものには移動物体 ID が与えられる．その際，異なる 2 つの移動物体の ID が与えられたブロックはいずれの移動物体からも除外する．すなわち ID を与えない．これにより，2 つの移動物体の混同を避けることができる．

5. 判定不能ブロックの取扱い

ステップ 4 において，画像上複数の移動物体が重なる場合には，一つのブロックに異なる移動物体の ID が複数与えられることになる．しかし，当該ブロックがいずれの移動物体に属するかを決定することはできないため，当該ブロックはいずれの移動物体の ID も与えられない．

6. 移動物体の分割

同一のスリット内を同時に通過する複数の移動物体が画像上で重なっているときなどは，単一の移動物体として生成される．この移動物体内で著しく動きベクトルが異なるクラスタや，背景画像で分離されるクラスタが存在するとき，クラスタごとに移動物体として分割される．

以上のアルゴリズムを適用し，車両をトラッキングしている様子を示す．図 4.3 において黒い矩形で囲われている部分が車両として追跡をしている部分である．

4.3 仮想空間への投影

トラッキングの結果は入力画像上の 2 次元の座標系における点として扱う．仮想空間において車両を適切な位置に配置するために，観測された車両の位置を実世界から仮想空間へ投影する．ここでは車両は 2 次元平面上に存在すると仮定する．これにより入力画像から仮想空間平面へのマッピングに関して 2 次元射影変換を適用することができる．2 次元射影変換は次のように表される．

$$\lambda \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (4.2)$$



図 4.2: 背景画像



図 4.3: トラッキングの結果

$$\begin{cases} x' = \frac{h_{11}x + h_{12}y + h_{13}}{h_{31}x + h_{32}y + h_{33}} \\ y' = \frac{h_{21}x + h_{22}y + h_{23}}{h_{31}x + h_{32}y + h_{33}} \end{cases} \quad (4.3)$$

射影行列 H を決定するために 4 点ずつの対応点を選び、 $h_{33} = 1$ として射影変換行列を求める。

この行列 H が実世界の平面から仮想空間への射影変換行列であり、 H^{-1} がその逆の仮想空間における平面から実世界の平面への変換行列となる。

4.4 視覚化

結果として、CG によって仮想空間の中を車両が走行する。これまでに述べた方法により、トラッキングされた結果は仮想空間における平面上の座標に投影される。それぞれのフレームごとに投影された座標を結び、その上を 3 次元モデルの車両が走る。

ユーザは車両が走る仮想空間を自由な視点から見ることができる。また視覚的な効果として前方にある建物を透かして表示させることができ、それらの背後を走る通常見えない車両を可視化することができる。実際の利用を考えた場合、これらは交差点などにおける出会い頭の事故防止支援のために役立つものと言える。

4.4.1 ガウシアンフィルタ

ブロックマッチングによる追跡結果の誤差などにより、フレーム間のトラッキングの結果をつないでも、車両の走る軌跡が平坦でないため CG によって表示を行った場合違和感が生じる。そこでここではトラッキングした点に対しガウシアンフィルタをかけることにより車両がスムーズに走行するようにした。

加重平均化は注目画素の距離に応じて異なる重み $w(k, l)$ をつけて加重平均濃度を出力する以下のような濃度変換を行う。

$$g(i, j) = \sum_{(k, l) \in N(i, j)} w(k, l) f(k, l) \quad (4.4)$$

ただし、 $\sum_{(k, l) \in N(i, j)} w(k, l) = 1$ になるように重み係数を正規化する。一般に、注目画素 (i, j) に近いほど大きな重みを設定し、出力値に対する寄与を大きくする。

重み $w(k, l)$ として 2 次元正規分布の確率密度関数 (2 次元ガウス関数)

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x^2 + y^2}{2\sigma^2}\right\} \quad (4.5)$$

を採用したものがよく用いられ，ガウシアンフィルタ (Gaussian filter) と呼ばれる．ここで標準偏差 σ^2 は平滑化の程度を制御するパラメータとなっている．

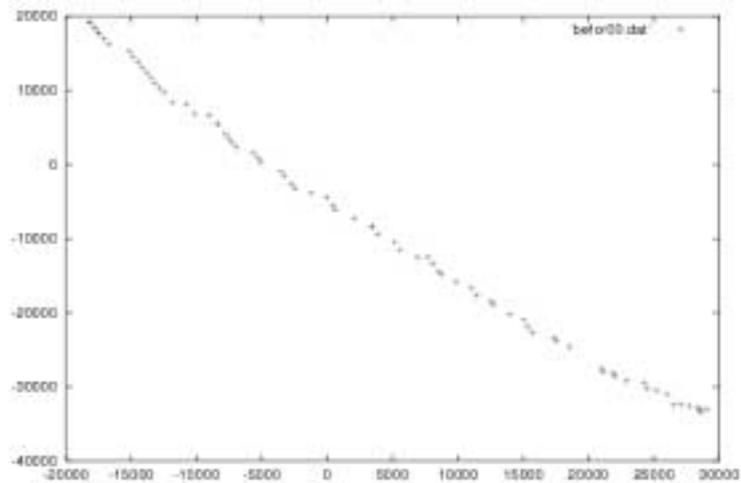
ここでは， σ の値を 3.0，フィルタのサイズを 51 として利用した．フィルタをかけない場合の車の追跡結果を表す点を図 4.4(a) に，またフィルタをかけた後の座標を図 4.4(b) に示す．これにより車両のスムーズな走行結果が得られた．

4.5 結果

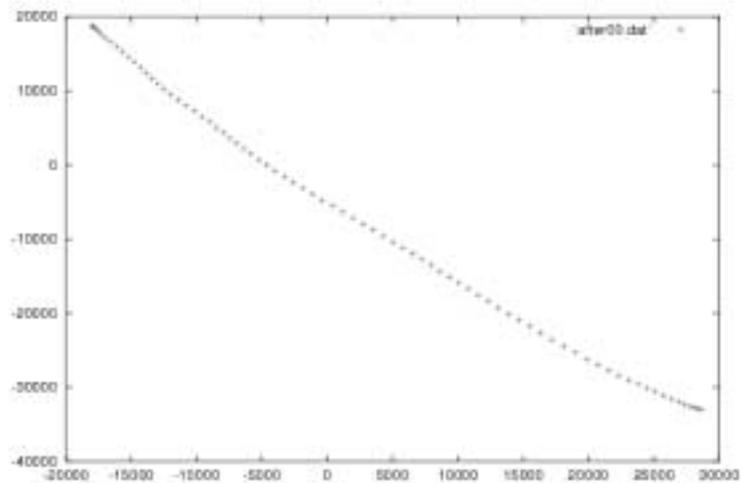
以上のステップにより得られた仮想空間の CG によるビューを示す．

まず図 4.5 にカメラから得られた入力画像を示す．これに対し，実世界にあるカメラと同じ位置，姿勢によりビューを再現したものが図 4.6 である．また同じ交通状況において，任意の視点からのビューを生成できる例として，図 4.7 を示す．最後に，運転席から眺めた様子を図 4.8 に示す．このうち，4.8(a) に関しては建物の壁面を透過させていることにより，左手より赤色の乗用車が交差点へ侵入しようとしているのがわかる．これに対し通常の運転では図 4.8(b) のような見えになり，侵入中の車を確認することはできない．

このように，任意視点から交差点での交通状況を視覚化し，他車の動きを確認できることはドライバの安全支援に有用であると思われる．



(a) フィルタリング前の軌跡



(b) ガウシアンフィルタリング後の軌跡

図 4.4: 車両の追跡結果を仮想空間座標系へ投影した結果



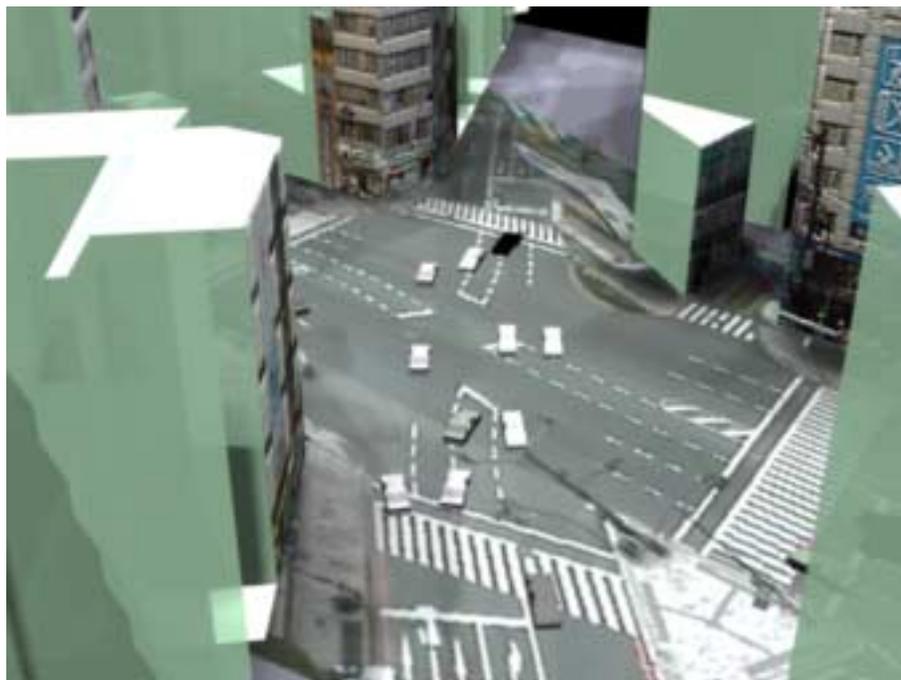
図 4.5: カメラから得られた入力画像



図 4.6: 実世界のカメラと同じ位置，姿勢を再現したビュー

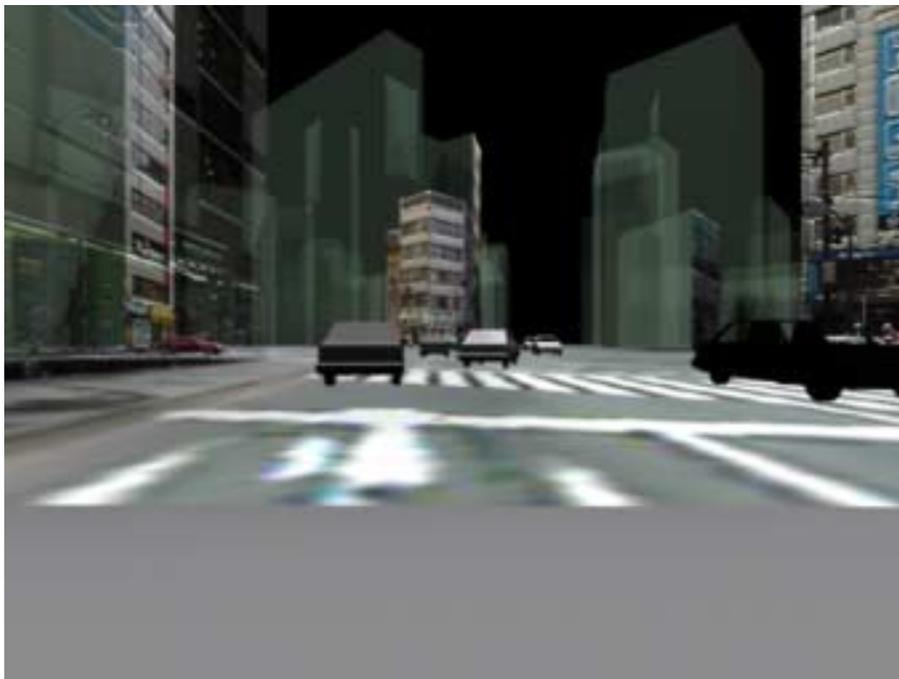


(a)

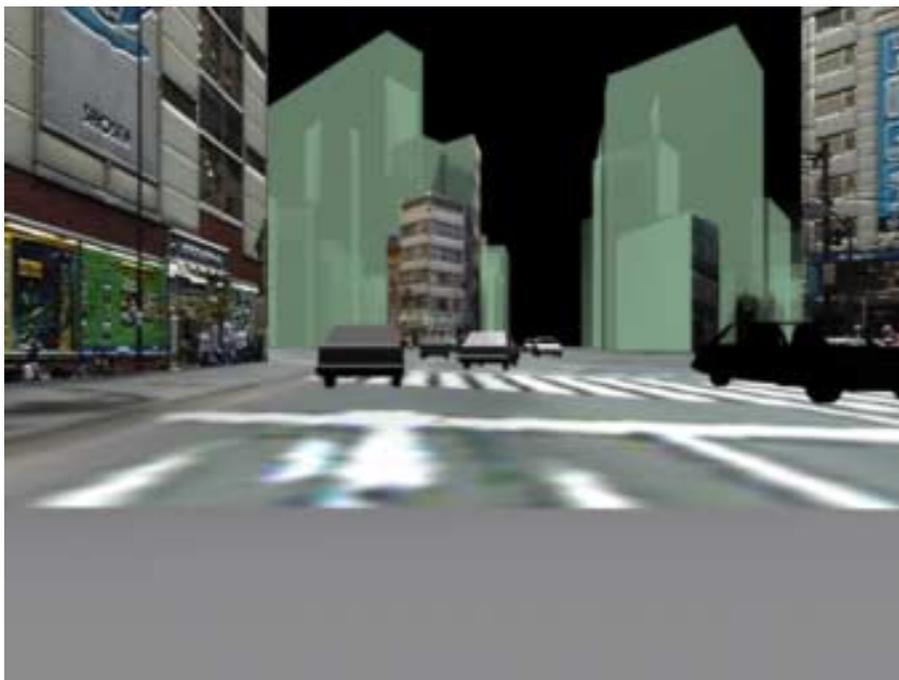


(b)

图 4.7: 鳥瞰映像



(a)



(b)

図 4.8: 運転席からの眺め

第5章 結論

本稿では、仮想空間に現実世界の情報を投影するという枠組を提案し、実際に ITS の主要な開発分野の一つである安全運転支援のための手段として、交通状況を視覚化しドライバに提供するシステムのプロトタイプシステムの実装を行った。実世界情報を仮想空間へ投影することを提案した。任意の視点から対象とするオブジェクトを含む空間を眺めることで実世界における対象オブジェクトとその周囲の環境との空間的な関係を容易に把握できるようになる。プロトタイプシステムの実装によりそのことが明らかになった。

しかし、このようなシステムを実際に ITS の分野に適用することを考えると、以下のような問題点が考えられる。

- 道路監視カメラの映像から車両を追跡し、それを CG により視覚化するという枠組では、車両のロバストな検出・追跡が求められる
- 実時間での検出・追跡処理が必要となる。
- 路車間通信におけるデータの配信プロトコルなどの検討が必要である。
- 表示のための 3 次元空間データの整備が必要である。

交差点のような環境においてに頑強な性能を持ったトラッキングは難しいといえる。本研究では仮想空間への情報の投影の一例として視覚情報をもとに投影を行ったが、他のセンサとの併用も考えられる。路車間通信や 3 次元データの整備に関しては現在それぞれ分野において活発な研究が行われており、今後の実現が望まれる。

参考文献

- [1] 入谷忠光, 宮武秀考, 片岡実, 山下悟郎, 大家隆弘, “見通し外地交差点の出会い頭事故防止にも有効な車車間通信方式,” 信学技報, SST99-46, ITS99-30, pp. 37–42, Nov. 1999.
- [2] E. Ichihara, H. Takao, Y. Ohta, “NaviView: Bird’s-Eye View for Highway Drivers Using Roadside Cameras,” Proc. of IEEE Int. Conf. on Multimedia Computing and Systems, pp.559–565, June 1999.
- [3] S. E. Chen, “QuickTime VR – an image-based approach to virtual environment navigation,” Proc. of ACM SIGGRAPH’95, pp. 29–38, 1995.
- [4] 星野准一, 増田功, “局所的な輝度変動にロバストなパノラマ画像生成法,” 電子情報通信学会論文誌, Vol. J82-D-II, No. 2, pp. 222–229, Feb. 1999.
- [5] 谷口行信, 阿久津明人, 外村佳伸, “PanoramaExcerpts: パノラマ画像の自動生成・レイアウトによる映像一覧,” 電子情報通信学会論文誌, Vol. J82-D-II, No. 3, pp. 390–398, Mar. 1999.
- [6] 千葉直樹, 蚊野浩, 美濃導彦, 安田昌司, “画像特徴に基づくイメージモザイク,” 電子情報通信学会論文誌, Vol. J82-D-II, No. 10, pp. 1581–1589, Oct. 1999.
- [7] R. Szeliski, “Video Mosaics for Virtual Environments,” IEEE Comput. Graphics & Appl., vol. 16, no. 3, pp. 22–30, Mar. 1996.
- [8] M. Yachida, M. Asada, S. Tsuji, “Automatic analysis of moving image,” IEEE Trans. Pattern Anal. Mach. Intell. vol. PAMI-3, no. 1, pp. 12–20, 1981.
- [9] 谷内田正彦, “動画像の処理と理解,” ロボットビジョン, 8章, 昭晃堂, 1990.
- [10] 佐藤淳, “コンピュータビジョン 視覚の幾何学,” コロナ社, 1999.
- [11] Collins, Lipton, Kanade, Fujiyoshi, Duggins, Tsin, Tolliver, Enomoto, and Hasegawa, “A System for Video Surveillance and Monitoring: VSAM Final Report,” Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May, 2000.

-
- [12] 須藤智, 小沢慎治, “サッカーのシーン解析,” 画像の認識・理解シンポジウム (MIRU2000) 講演論文集 II, pp. II-223 – 228, 2000.
- [13] 高井勇志, 土井元紀, 千原國宏, “映像解析と VR 技術を用いたフットボール練習支援システム,” 日本バーチャルリアリティ学会第 4 回大会論文集, pp. 233–234, 1999.
- [14] S. Kamijo, Y. Matsushita, K. Ikeuchi, M. Sakauchi, “Traffic Monitoring and Accident Detection at Intersections,” IEEE Trans. ITS, vol. 1, No. 2, pp. 108–118, June 2000.
- [15] K. Kamei, W. Hoy, T. Tamada, K. Seo, “Modeling of Urban Scenes by Aerial Photographs and Simply Reconstructed Buildings,” Proc. of MVA’98, Nov 1998, pp.224–227.
- [16] O. Faugeras, L. Robert, S. Laveau, G. Csurka, C. Zeller, C. Gauclin, I. Zoghلامي, “3-D Reconstruction of Urban Scenes from Image Sequences,” COMPUTER VISION AND IMAGE UNDERSTANDING, Vol. 69, NO. 3, Mar 1998, pp.292–309.
- [17] Y. Ogawa, K. Iwamura, S. Kakumoto, “A Map-Based Approach for Extracting Object Information from Aerial Images,” Proc. of IAPR Workshop on Machine Vision Applications (MVA’98), Nov 1998, pp.220–223.
- [18] H. Zhao, R. Shibasaki, “A System for Reconstructing Urban 3D Objects using Ground-based Range and CCD Sensors,” Proc. of International Workshop on Urban 3D/Multi-Media Mapping (UM3’99), Sept 1999, pp.175–182.