

NAIST-IS-MT9551033

## 修士論文

# 段階的なモデルマッチングによる動画像における 歩行者の追跡と映像合成への応用

草地 良規

1997年2月14日

奈良先端科学技術大学院大学  
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に  
修士(工学) 授与の要件として提出した修士論文である。

草地 良規

指導教官： 横矢 直和 教授  
千原 國宏 教授  
竹村 治雄 助教授

# 段階的なモデルマッチングによる動画像における 歩行者の追跡と映像合成への応用\*

草地 良規

## 内容梗概

現実には存在しないシーンや撮影が困難なシーンを映像合成により創るには、創作された映像の中の物体間の隠蔽関係や床と人物の接地関係など、物体間の合理的な位置関係の保持が必要となる。動画像中の人物を他の人物に入れ換えると、合理的な位置関係を保った映像合成が可能である。そのためには、動画像中の人物の動き情報の抽出と入れ換える人物の3次元CGモデルが必要である。

本論文では、動画像中の人物を入れ換える映像合成手法の基礎として、固定カメラで撮影された動画像中の、直線的に歩行する人物を入れ換える映像合成手法を扱う。まず、動画像中の人物の動き情報を抽出するために段階的なモデルマッチングを提案する。本手法は、人間の3次元モデルのパラメータを段階的に求めて探索の計算コストを削減し、脚の交差などによる自己隠蔽問題を解決する手法である。次に、人体を関節物体とみなして、人物の3次元CGモデルを構築する手法を提案する。最後に、実画像に対し歩行者追跡及び映像合成実験を行ない、合理的な位置関係を保った映像合成が可能であることを示す。

## キーワード

モデルマッチング, 歩行者追跡, 映像合成, 関節物体, 3次元CGモデル, 自己隠蔽

---

\*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 修士論文, NAIST-IS-MT9551033, 1997年2月14日.

# Tracking of a Walking Person in Image Sequences by Stepwise Model Matching and Its Application to Image Synthesis\*

Yoshinori Kusachi

## Abstract

Image synthesis can create scenes which are not in existence or are difficult to be taken. It is necessary to locate objects in created images in a consistent manner considering occlusion among objects. If, for example, a person in image sequences can be replaced by another person, it is easy to keep reasonable positional relationship between objects.

In this thesis, for establishing the basis of such an image synthesis, I propose a method of substituting a walker with 3-D CG model of another person. The person is assumed to walk straightforward in an image sequence taken by a fixed camera. Motion parameters of the walker in the image sequence are obtained by tracking the walker using a stepwise model matching technique. This method can reduce the cost of obtaining motion parameters of human 3-D model and can track self-occluding parts. 3-D CG model of a person is constructed, regarding a person as an articulated object. Experimental results show that the proposed method can achieve an image synthesis.

## Keywords:

model matching, tracking of a walker, articulated object, image synthesis, 3-D CG model, self occlusion

---

\*Master's Thesis, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-MT9551033, February 14, 1997.

# 目次

1. はじめに	1
2. 従来の映像合成技術と本研究の方針	3
2.1 従来の映像合成手法	3
2.1.1 被写体を背景画像に重ねて描く映像合成手法	3
2.1.2 被写体を入れ換える映像合成手法	5
2.1.3 要素技術に関する考察	7
2.2 本研究の方針	9
2.2.1 処理の概要	10
2.2.2 本研究の特徴	12
3. 段階的なモデルマッチングを用いた歩行者の追跡	13
3.1 動画像からの人物領域の抽出	14
3.2 標準人体モデル	15
3.3 評価関数の設定によるモデルマッチングの定式化	16
3.4 3次元モデルマッチングの問題点と解決法	18
3.5 歩行の制約条件の導入	20
3.5.1 定性的な制約条件	20
3.5.2 定量的な制約条件	20
3.6 段階的なモデルマッチング	22
3.6.1 全体探索空間と部分探索空間	22
3.6.2 パラメータの依存関係と運動の独立性	25
4. 歩行者の追跡結果を用いた映像合成	32
4.1 人物の3次元CGモデルの構築	32
4.2 人物の3次元CGモデルのアニメーション手法	36
4.3 静止物体と人物に隠蔽関係がある場合のマスク画像の作成	37
5. 実験	41

5.1	実験の環境と設定 . . . . .	41
5.2	実験 1 :横に歩行する人物の追跡及び映像合成 . . . . .	41
5.3	実験 2 :斜めに歩行する人物の追跡及び映像合成 . . . . .	48
5.4	実験 3 :物体に隠蔽される歩行者の追跡及び映像合成 . . . . .	53
6.	おわりに	58
	謝辞	60
	参考文献	61

## 目 次

2.1	歩行者を入れ換える映像合成の処理の流れ . . . . .	11
3.1	背景抽出の手法 . . . . .	14
3.2	標準人体モデルとパラメータ . . . . .	16
3.3	$E_{model}$ の定義 . . . . .	18
3.4	遊脚の位置関数 . . . . .	22
3.5	全体探索空間と部分探索空間 . . . . .	23
3.6	段階的なモデルマッチングの概念図 . . . . .	24
3.7	構成した部分探索空間 . . . . .	27
3.8	パラメータの支配・決定関係 . . . . .	28
3.9	段階的なモデルマッチング . . . . .	29
4.1	獲得したぬいぐるみの3次元CGモデル . . . . .	34
4.2	頭, 胴体の3次元CGモデリングのための実物体の計測 . . . . .	35
4.3	腕, 脚の3次元CGモデリングのための実物体の計測 . . . . .	36
4.4	脚部の正規化 . . . . .	37
4.5	マスク画像の作成 . . . . .	39
5.1	原画像 (横歩行) . . . . .	44
5.2	背景画像 (横歩行) . . . . .	45
5.3	人物領域の抽出及び歩行者の追跡結果 (横歩行) . . . . .	46
5.4	追跡結果に基づくCGモデルの合成 (横歩行) . . . . .	47
5.5	原画像 (斜め歩行) . . . . .	49
5.6	背景画像 (斜め歩行) . . . . .	50
5.7	人物領域の抽出及び歩行者の追跡結果 (斜め歩行) . . . . .	51
5.8	追跡結果に基づくCGモデルの合成 (斜め歩行) . . . . .	52
5.9	原画像 (物体による隠蔽) . . . . .	54
5.10	背景画像 (物体による隠蔽) . . . . .	55
5.11	マスク画像 (物体による隠蔽) . . . . .	55
5.12	人物領域の抽出及び歩行者の追跡結果 (物体による隠蔽) . . . . .	56
5.13	追跡結果に基づくCGモデルの合成 (物体による隠蔽) . . . . .	57

## 表 目 次

2.1	従来の映像合成手法の利点と欠点 . . . . .	6
2.2	入れ換える映像合成の利点と欠点 . . . . .	7
5.1	実験パラメータ . . . . .	42



## 1. はじめに

近年、映像合成に対する要求は多様化しており、現実には存在しないシーンや撮影が困難なシーンの映像を創りたいという要求が特に高い。しかし、このような映像合成では、物体同士の隠蔽関係や床と物体の接地関係が矛盾しないことを必要とし、実現に手間がかかる。

位置関係が矛盾しない合成映像を得るために、仮想スタジオの手法が提案されている [1]。この手法は、あらかじめ位置関係を考慮して合成素材（映像合成に使用される映像）を用意し、合成素材を重ねて描く手法である。しかし、この手法では、4 方の壁の色を特定するなど撮影環境を制限しているため、テレビ番組や映画などの既存の映像資産を合成素材とするのは困難となる。また、仮想スタジオには他にも問題があり、創作可能な合成映像が制限されてしまう。

これに対し、動画像中の物体や人物を他のデータに入れ換えるという映像合成手法を考える。このとき被写体の領域と 3 次元の位置情報を画像認識の技術を用いて動画像から抽出できれば、3 次元の位置関係が矛盾しない合成映像を得ることができる。動画像中の物体を入れ換える映像合成は、撮影環境を制限しないため、合成素材として既存の映像資産を利用できる。また、撮影後に不要な物体を消去したり役者を変更する修正や編集も可能であり、自由に映像を創造できる。この映像合成手法の中で、静止物体を入れ換える手法は、すでに提案されている [2]。しかし、この手法では動物体を入れ換えることはできない。そのため、動物体を、特に人物を入れ換える映像合成手法の確立が望まれる。

動画像中の移動人物像を、位置関係が矛盾しないように他の人物像に入れ換えるためには、動画像中の正確な移動人物領域を抽出することと、人物の動き情報（人物の関節角などの、人物の位置と姿勢を決定する情報）を得る必要がある。次に、これらの情報より他の人物像系列を得る。他の人物像系列を得るには、他の人物の 3 次元 CG モデルを作成し、動き情報を基に人物の 3 次元 CG モデルを動かす、アニメーションを作成する必要がある。

本研究では、人物を入れ換える映像合成手法の基礎として、固定カメラで撮影された動画像中の、直線的に歩行する人物を入れ換える映像合成手法の開発を目的とする。歩行は人間の基本的な動作であり、この手法は種々の動作をする人物を

入れ換える映像合成に応用できる。

本研究では人物の動き情報を得る手法として、段階的なモデルマッチングにより移動人物の追跡を行なう手法を提案する。動き情報を得るために多くの研究がなされているが、人間の動きは自由度が高いため、完全な動き情報を得る手法は確立されていない。そのため、人間の動きを基本的な動作である歩行などに限って動き情報を得る研究がある [3, 4, 5]。しかし、これらの手法では、脚の交差による自己隠蔽問題を完全には解決できない。また、動き情報の誤差は避けられず、生成した動作が不自然になる。一方、本研究における段階的なモデルマッチングは、歩行者の肘, 膝, 肩, 股の関節パラメータと頭の位置パラメータを段階的に求めて探索の計算コストを削減し、自己隠蔽問題を解決する。その際、歩行の特徴をとらえた運動パラメータを抽出することにより、動き情報に誤差があっても自然なアニメーションとなる。

次に、人物の3次元CGモデルの獲得手法として部位ごとに計測する手法を提案する。物体の3次元CGモデルは手作業で入力又は構築を行なっているのが現状である。しかし、一般に複雑な物体の3次元CGモデルは手作業による構築が困難であるため、レンジファインダなどを用いて物体を計測し、3次元CGモデルを構築する研究が行なわれている [6]。ところが、従来手法では関節物体を考慮していないため、人物の3次元CGモデルの獲得には不向きである。本研究では人物を関節物体とみなして各部位ごとに計測を行ない人物の3次元モデルを構築する。

以下、2章で本研究の位置づけと方針を述べ、3章で提案手法である段階的なモデルマッチングの手法を述べる。また、4章で歩行者の追跡結果を用いた映像合成を述べる。5章では実画像に本手法を適用した実験結果及び考察を述べ、6章でまとめと今後の課題を述べる。

## 2. 従来の映像合成技術と本研究の方針

近年, 映像合成に対する要求は多様化しており, 特に現実には存在しないシーンや撮影が困難なシーンの映像を創りたいという要求が高い. こうした要求には次のようなものがある.

- 撮影が不可能な映像を創りたいという要求.

具体例 : テレビや映画などにおいて, 過去や未来, 異次元世界などの映像や, 既に死んでいる俳優が出演する映像などを創りたいという要求.

- 撮影が困難な映像を創りたいという要求.

具体例 : 人が車にはねられるシーンの映像などを創りたいという要求.

- 撮影後に映像を修正したり変更したりしたいという要求.

具体例 : 映像の誤りが撮影後に見つかった場合に修正を加えたり, 出演者の怪我などにより後の撮影ができない場合, それまでに撮影された映像中の出演者を他の人物に入れ換えたいという要求.

これらの映像合成を手作業で行なうと, 見ために自然でリアルな映像を創り出すことができる. しかし, 計算機上で1フレーム毎に映像を作成するため, 特化した専門家の極めて根気強く丹念な作業を要し, 完成までに時間がかかる. そのため, これらの映像合成を自動で行なうことのできる映像合成手法の確立が望まれる.

本章では, 2.1 節で従来の映像合成手法についての考察を, 2.2 節で本研究の方針を述べる.

### 2.1 従来の映像合成手法

#### 2.1.1 被写体を背景画像に重ねて描く映像合成手法

映像合成を自動で行なうために, 合成素材 (映像合成に使用される映像) を重ねて描く映像合成手法が実用化されている. 以下にその手法を説明する.

- クロマキー合成

クロマキーブルーと呼ばれる特殊な背景色をバックに物体を撮影した映像から、クロマキーブルーの領域を排除することにより物体を切り出して、背景画像に重ねて描く手法である。古くから確立されている手法であり、処理が高速（リアルタイムで可能）であるという利点がある。しかし、次のような問題点がある。

- クロマキーブルーが一様な照明環境下で撮影されなければならない、
- カメラの移動が禁止される。
- 合成素材間の隠蔽関係が強く制限される。
- 撮影後の修正が困難である。
- 照明器具やクロマキーブルーの背景などを必要とし、システムが大規模である。

- 仮想スタジオ

近年の CG 技術の発展に伴い実用化された、クロマキー合成にカメラの移動を考慮した手法である [1]。仮想スタジオでは、まず、4 方がクロマキーブルーであるスタジオを用意し、物体をクレーンカメラで撮影する。撮影された映像からクロマキーブルーの領域を排除し、被写体領域を得る。撮影の際、クレーンカメラの移動（パン、チルト、ズーム、平行移動）パラメータを記録しておく。次に、そのパラメータに従って仮想空間の見え方を変化させて CG 画像を得る。切り出した被写体領域を CG 画像に重ねて描くことにより映像合成を行なう。この手法には、合成素材をライブラリ化することで大道具やロケを不要にできる、映像合成の処理が高速である、などの利点がある。さらに、背景の床面と人物の接地関係が考慮されており、幾何学的に矛盾しない映像合成が可能である。しかし、次のような問題がある。

- 合成素材間の矛盾のない隠蔽関係を保つことが困難である。
- 撮影後の修正が困難である。

- 照明器具や4方がクロマキーブルーの背景などを必要とし、システムが大規模である。

実用化されているこれらの手法では、ここに挙げたいくつかの問題点により、創造できる映像が制限されている。そこで、自由にかつ手軽に映像を創造できる映像合成手法が望まれる。まず、自由な映像合成を行なうためには、以下の点を考慮する必要がある。

- 合成素材間の接地関係が矛盾しない（接地関係）
- 合成素材間の隠蔽関係が矛盾しない（隠蔽関係）
- 撮影環境が限定されない（自由な合成素材）

次に、手軽に映像合成を行なうためには、以下の点を考慮する必要がある。

- 撮影後の修正が可能である（修正可能）
- 人間の行なう処理が少ない（自動化）
- システムが小規模であり、システムの設置が容易である（設置の容易さ）
- 合成処理が単純で時間がかからない（処理の実時間性）

これらの点に関して表 2.1 に従来の映像合成手法の特徴をまとめる。

自動化されている映像合成手法の問題点を解決するために、画像認識の技術を用いて自動的に動画像中の被写体を他の物体に入れ換える映像合成が考えられる。

### 2.1.2 被写体を入れ換える映像合成手法

被写体を入れ換える映像合成とは、動画像中の被写体の領域を切り出し、他の物体（3次元CG）に入れ換えるという合成手法である。動画像中の被写体の領域と3次元の位置情報を、画像認識の技術を用いて自動的に抽出することができれば、動画像中の被写体を他の物体に入れ換えることにより、正しい接地関係と隠蔽関係を保った映像合成を行なうことができる。また、撮影環境も限定されず、既存

表 2.1 従来の映像合成手法の利点と欠点

評価項目	クロマキー合成	仮想スタジオ	手作業
接地関係	○	○	○
隠蔽関係	△	△	○
自由な合成素材	×	×	○
修正可能	×	×	○
自動化	○	○	×
設置の容易さ	×	×	○
処理の実時間性	○	△	×

○…考慮されている, ×…考慮されていない, △…考慮されているが難しい

の映像資産にも適用でき, さらに, 撮影後の修正も可能である. しかし, 動画像中の被写体の領域を抽出することや, 被写体の領域から被写体の3次元の位置情報を獲得するためには複雑な処理を必要とし, 映像合成処理に時間がかかるという問題点がある. 表 2.2 に被写体を入れ換える映像合成の利点と欠点をまとめる.

表 2.1 と表 2.2 を比較すると分かるように, 被写体を入れ換える映像合成が実現できれば, 処理は複雑ではあるが, 実用化されている手法において考慮されていない問題を克服できる可能性がある. 被写体を入れ換える映像合成手法の具体的な処理手順は以下のようなになる.

- (1). 被写体領域と背景領域を作成する.
- (2). 被写体領域から被写体の3次元の位置情報を抽出する.
- (3). 入れ換える物体の3次元 CG モデルを作成する.
- (4). 3次元の位置情報を基にアニメーションを作成する.
- (5). アニメーションを被写体領域に重ねて描く.

被写体を入れ換える映像合成の例として, 動画像中の看板を他の看板 (2次元 CG) に入れ換えるという研究がある [2]. しかし, この手法では入れ換える物体を

表 2.2 入れ換える映像合成の利点と欠点

評価項目	入れ換える映像合成
接地関係	○
隠蔽関係	○
自由な合成素材	○
修正可能	○
自動化	○
設置の容易さ	○
処理の実時間性	×

○…考慮されている, ×…考慮されていない, △…考慮されているが難しい

平面的な静止物体と仮定しているため, 動物体を考慮していない. もし動物体を入れ換えることができれば, 例えば主人公の入れ換えや模型と実物の入れ換えなどの映像合成が可能になる.

動物体を入れ換える映像合成の中で最も要求が高いのは, 人物を入れ換える映像合成である. 人物を入れ換える場合, 映像合成に必要な情報は,

- (i). 移動人物領域
- (ii). 人物の動き情報
- (iii). 入れ換える人物の 3 次元 CG モデル

である. つまり, 本手法の実現に必要となる要素技術は移動人物領域の抽出手法, 動き情報の抽出手法, および人物の 3 次元 CG モデルの作成手法である. 次節ではこれらについて考察する.

### 2.1.3 要素技術に関する考察

はじめに, 前節 (i) の移動人物領域の抽出について考える. 移動物体領域の抽出はロボットビジョンなどの分野で盛んに研究が行なわれている. 移動物体領域を

求める代表的な手法としては、カメラの移動パラメータを求め、そのパラメータを基に背景領域を抽出し、その他の領域を移動物体領域とする手法がある [7, 8]. しかし、人間はいつも体全体を動かしている訳ではないため、この手法では人物領域を正確に求めることはできない. 例えば人間の歩行では片方の脚下は静止しているため、この手法では脚下が背景領域として抽出されてしまう. Niyogi らはカメラを固定として、歩行者領域を正確に求めているが [3], 人間の膝の高さを既知としている.

次に、前節 (ii) の動き情報の抽出について考える. 動き情報の抽出は、画像認識の分野で盛んに研究されている. 代表的な手法に、人物領域の心線化による動作解析手法など、画像特徴から動き情報を求めるボトムアップ的な手法がある [9]. しかし、ボトムアップ的な手法はモデルを持たないため、3次元の動き情報を得るのは困難である. そのため、近年、物体のモデルをあらかじめ保持し、画像特徴とモデルの整合性を評価する3次元モデルマッチングなどのトップダウン的な手法の研究が盛んである [4, 5, 10, 11, 12, 13].

移動人物は関節物体と見なすことができる. そのため、動き情報の獲得は、連続したフレームでの関節物体の姿勢推定と追跡と考えることができる. Lowe は3次元モデルマッチングによって関節物体の姿勢推定を行なっている [11]. しかし、移動人物などの複雑な関節物体であると内部パラメータが多くなり、探索空間が膨大になるという問題点がある. 亀田らは、各部位ごとにパラメータを求めることで、一般的な関節物体の姿勢推定を行なっているが [12], 関節物体の正確なモデルを必要としているため、個人によってモデルの異なる人物の追跡には不向きである.

そこで、詳細なモデルを持たず、円筒などのプリミティブで構築したモデルを用いてモデルマッチングにより動き情報を得る研究が行なわれている. 山本らはオプティカルフローを評価して3次元モデルマッチングを行ない、人物の追跡を行なっている [4]. また、持丸らは人物のテクスチャ情報を評価して2次元モデルマッチングを行ない、人物の動作解析を行なっている [5]. しかし、これらの手法は、脚の交差などによる自己隠蔽問題を避けるため、全身像を追跡していない. そのため、動き情報を完全に得ることはできない. 島田らは、指の姿勢の事前確率を



導入して、3次元モデルマッチングにより自己隠蔽を考慮した手指の姿勢推定を行なっている [13]. しかし人間のモデルは左右対象であるため、人間の全身像を対象にした場合、確率によって自己隠蔽問題を解決することはできない。つまり、全身の人物像を追跡しつつ自己隠蔽問題を解決する手法は確立されていない。

そこで、人間の動きに対し制約を導入して動き情報を得る研究が行なわれている。例えば、人物の動きを基本的な動きである直線的な歩行に限定している。Niyogiらは、時空間画像を用いてカメラの光軸に対し垂直方向に歩行する人物の動き情報を得ている [3]. また、亀田らは慣性を利用して自己隠蔽問題を解決している [14]. また、Polanaらは、テンプレートマッチングにより、歩行を含む周期的な動きを認識するシステムを構築している [15]. しかし、これらの手法では、動き情報の誤差は避けられない。誤差を含んだ動き情報に基づいて生成した動作アニメーションは不自然な動きとなる。

最後に、前節 (iii) の人物の3次元CGモデルについて考える。従来では、物体の3次元CGモデルは、人間の手作業によって構築されている。しかし、複雑な人物の3次元CGモデルの入力には多大な時間と労力が必要である。このような背景から、レンジファインダなどで実物体を計測して3次元CGモデルを構築する研究が盛んになってきている [6, 16, 17]. しかし、従来の手法では、関節物体を考慮していないため、歩行アニメーションの作成には不向きである。

## 2.2 本研究の方針

本研究では、人物を入れ換える映像合成の基礎として、カメラ固定の環境下で撮影された動画像中の、直線的に歩行する人物の入れ換え映像合成手法の開発を目的とする。

2.1.2節で述べた (i) の移動人物領域は、背景画像を作成し、動画像との差分を計算することで人物領域を獲得する。また、(ii) の動き情報の抽出手法として、本研究では、歩行者の全身像を追跡する段階的なモデルマッチングを提案する。段階的なモデルマッチングは、歩行者の頭の位置、膝、股、肩、肘の関節パラメータを段階的に求めることにより、探索の計算コストの削減を達成する。また、歩行の場合、左右の脚は交互に前に出るとする、どちらかの脚は軸脚（接地している脚）でその

他の脚は遊脚（軸脚に変わる瞬間以外は接地していない脚）とするなどの定性的な制約条件を導入することにより自己隠蔽問題を解決する。さらに、歩行に対し定量的な制約条件を導入することにより、動き情報に誤差があっても歩行アニメーションが自然であることを目指す。歩行の定量的な制約条件は、腕及び遊脚（軸脚に変わる瞬間以外は接地していない脚）の動きが解析学的、動力学的に矛盾のない歩行関数に従うとする条件である。人間の歩行アニメーションはよく研究されている分野であり、単純な歩行アニメーションを基に、階段や坂の登り下りの動作、荷物を持った歩行、感情を持った歩行、などのアニメーションに拡張が可能である [18, 19, 20, 21, 22]。そのため、本研究で提案する歩行者追跡手法もそのような動作に適用できるように拡張可能であると考えられる。(iii) の人物の 3 次元 CG モデルは部位ごとの計測により人物の 3 次元 CG モデルを構築する手法を提案する。

### 2.2.1 処理の概要

歩行者を入れ換える映像合成の処理の流れを図 2.1 に示す。まず、背景画像を作成し動画像のフレーム全体の人物領域を求める。次に、3 次元モデルマッチングに用いる標準人体モデルを作成する。さらに、動き情報を求めるために段階的なモデルマッチングを行なう。これらの処理の詳細を 3 章で述べる。

次に、映像合成の合成素材を準備する。まず、入れ換える人物の 3 次元 CG モデルを構築する。また、人物が静止物体に隠蔽される場合には、人物よりも手前にある領域を抽出しマスク画像を作成する。最後に合成映像を作成するために、次の処理を各フレームに対して行なう。

1. 背景画像を描画する。
2. 動き情報に従い、入れ換える人物の 3 次元 CG モデルの姿勢を決定する。
3. 入れ換える人物の 3 次元 CG モデルを上描きする。
4. 人物が静止物体に隠蔽される場合には、マスク画像を上描きする。

これらの処理の詳細を 4 章で述べる。

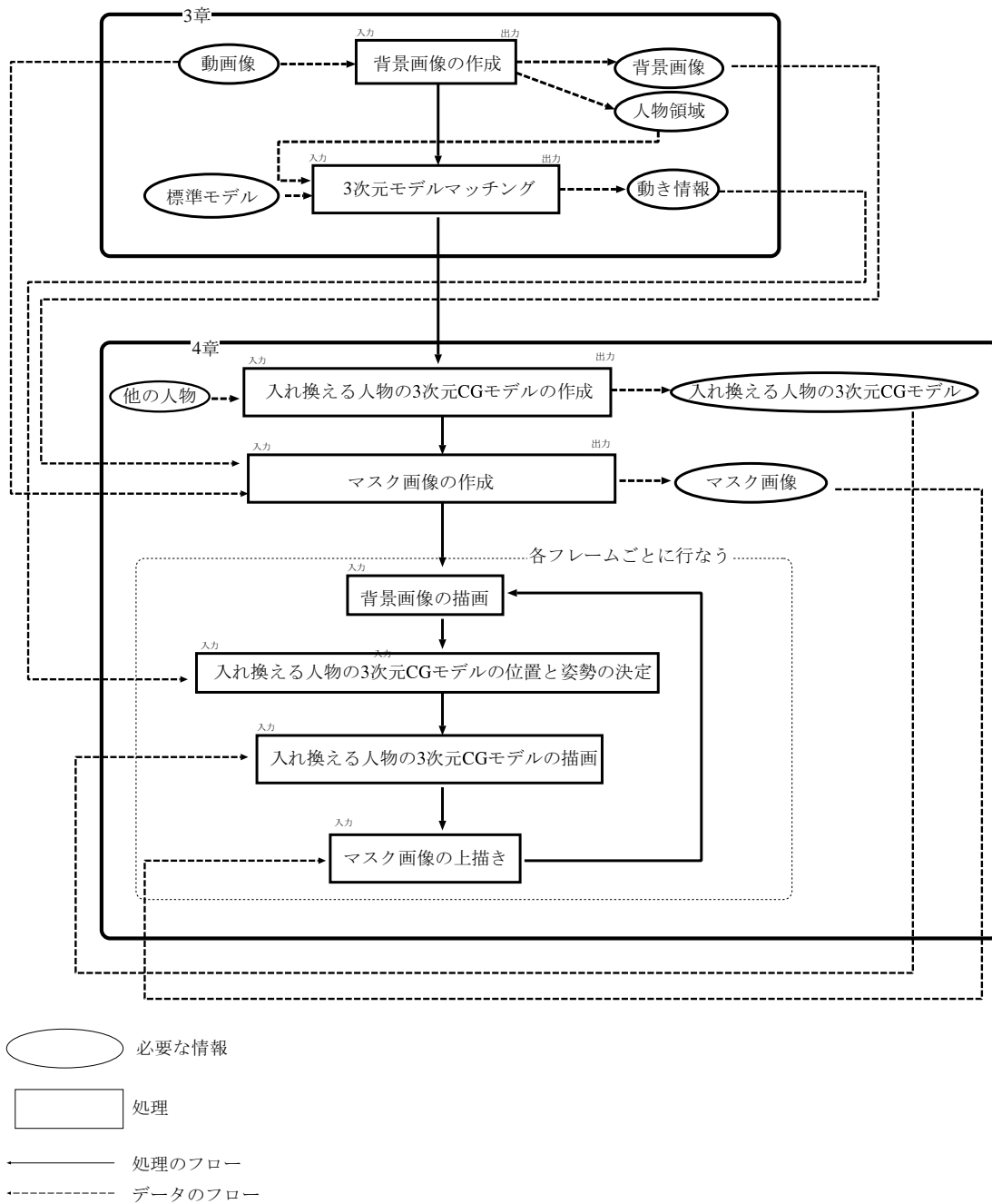


図 2.1 歩行者を入れ換える映像合成の処理の流れ

### 2.2.2 本研究の特徴

本研究の提案手法とその特徴を以下にまとめる。本研究では、動画像中の直線的に歩行する人物を他の人物の3次元CGモデルに入れ換える映像合成手法を提案する。その特徴として以下の点が挙げられる。

- 合理的な位置関係を保つことが可能である。
- 撮影後の修正が容易である。
- 合成素材を限定しない。

また、この手法の要素技術として、段階的なモデルマッチングを用いた歩行者の追跡手法と、関節物体を考慮した3次元CGモデリング手法を提案する。前者の特徴としては、探索における計算コストの削減、自己隠蔽問題の解決、歩行の特徴をとらえた運動パラメータの抽出が挙げられる。また、後者の特徴としては、計測によるリアルな3次元CGモデルの作成が挙げられる。

### 3. 段階的なモデルマッチングを用いた歩行者の追跡

本章では, 歩行者の追跡手法について述べる. まず, 背景画像を作成することで移動人物領域を抽出する. 次に3次元モデルマッチングにより, 歩行者を追跡し, 動き情報を得る. 3次元モデルマッチングを行なうためには

- 標準人体モデルの定義

関節角や3次元位置のパラメータを持つ人間の3次元モデル (標準人体モデル) の定義.

- 問題の定式化

人物領域と標準人体モデルの照合により標準人体モデルのパラメータを決定するための定式化

を行なう必要がある. 本研究では標準人体モデルを球, 円筒, 楕円筒のプリミティブで構成する. また, 人物領域と標準人体モデルの投影領域が重なる面積をモデルマッチングの照合度として, 問題の定式化を行なう.

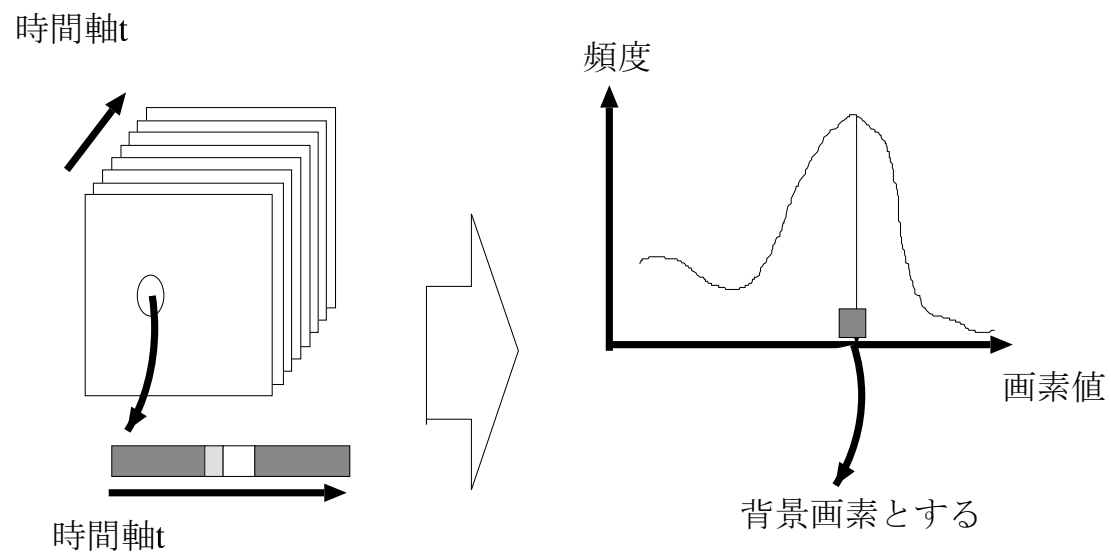
3次元モデルマッチングを用いて歩行者追跡を行なう場合, 自己隠蔽問題と誤差の問題がある. 本研究では歩行の際, 左右の脚は交互に前に出るとする, どちらかの脚は軸脚で一方の脚は遊脚とする, などの定性的な制約条件を導入することにより自己隠蔽問題を解決する. さらに, 腕及び遊脚の関節角を解析学的かつ動力学的に矛盾のない周期的な歩行関数で近似できるとする定量的な制約条件を導入することにより, 歩行の特徴をとらえた運動パラメータを抽出する.

最後に, 歩行者の胴体の位置, 膝, 股, 肩, 肘の関節パラメータを段階的に求めることにより, 3次元モデルマッチングの計算コストを削減する段階的なモデルマッチングを提案する.

まず, 3.1節で, 移動人物領域の抽出手法について述べる. 次に, 3.2節で標準人体モデルについて述べ, 3.3節でモデルマッチングの問題の定式化を示す. さらに, 3.4節で3次元モデルマッチングによる歩行者追跡の問題点について述べ, 3.5節で歩行の制約条件について述べる. 最後に, 3.6節で段階的なモデルマッチングについて詳しく述べる.

### 3.1 動画像からの人物領域の抽出

2章で述べたように、本研究ではカメラを固定とする。動画像において、ある位置の画素に対して時系列の画素値のヒストグラムを作成し、ピーク値を背景画素とすることで背景画像を作成する（図 3.1）。次に動画像の画素値と背景画像画素値の差分が閾値  $T_d$  以下の画素を人物領域画素とすることで人物領域の抽出を行う。



各画素でヒストグラム作成

図 3.1 背景抽出の手法

しかし、本手法では人物領域を正確に抽出することは不可能である。その原因は2つ考えられる。

- 背景と人物領域が同じ色をしている画素が存在すること
- 人物が長時間撮像される画素が存在すること

両画素ともに背景画素として認識されてしまう。

これらの原因により抽出した人物領域にはノイズが多く含まれてしまう。この不完全な人物領域から人物の動き情報を求める必要がある。そのため、本研究ではノイズに強いと考えられる3次元モデルマッチングの手法を用いる。

## 3.2 標準人体モデル

人間は個人によって体形が異なるため、汎用性の求められる歩行者の追跡においては、人間の正確なモデルを持つことができない。そのため、本研究では、3次元モデルマッチングで使用する標準人体モデルを、頭を球で、首、上腕、下腕、上脚、下脚を円筒、また、胴体を楕円筒で構成する。人間が一般的な動きをする場合、腕や脚は図 3.2 B) のように関節を原点としてオブジェクト座標系の 3 軸 (歩行の進行方向を  $z_{loc}$  軸, 進行方向垂直な方向を  $x_{loc}$  軸, 円筒の中心軸を  $y_{loc}$  軸) 周りの回転運動を行なう。

本研究では、人物が直線的な歩行をする際、 $x_{loc}$  軸周りに脚や腕を大きく回転するため、図 3.2 A) のように、ある時刻  $t$  のパラメータを、頭の位置 ( $z[t] = q_0[t]$ ), 左右肩の角度 ( $q_1[t], q_2[t]$ ), 左右肘の角度 ( $q_3[t], q_4[t]$ ), 左右股の角度 ( $q_5[t], q_6[t]$ ), 左右膝の角度 ( $q_7[t], q_8[t]$ ) の 9 個とする。この 9 個  $\times$  (動画像のフレーム数) のパラメータが動き情報になる。実際には人間の動きは自由度が高いため、人間は直線的な歩行を行なう際でも、 $z_{loc}$  軸,  $y_{loc}$  軸周りに脚や腕を回転する。しかし、 $x_{loc}$  軸周りの脚や腕の回転に比べて無視できるくらい小さい。

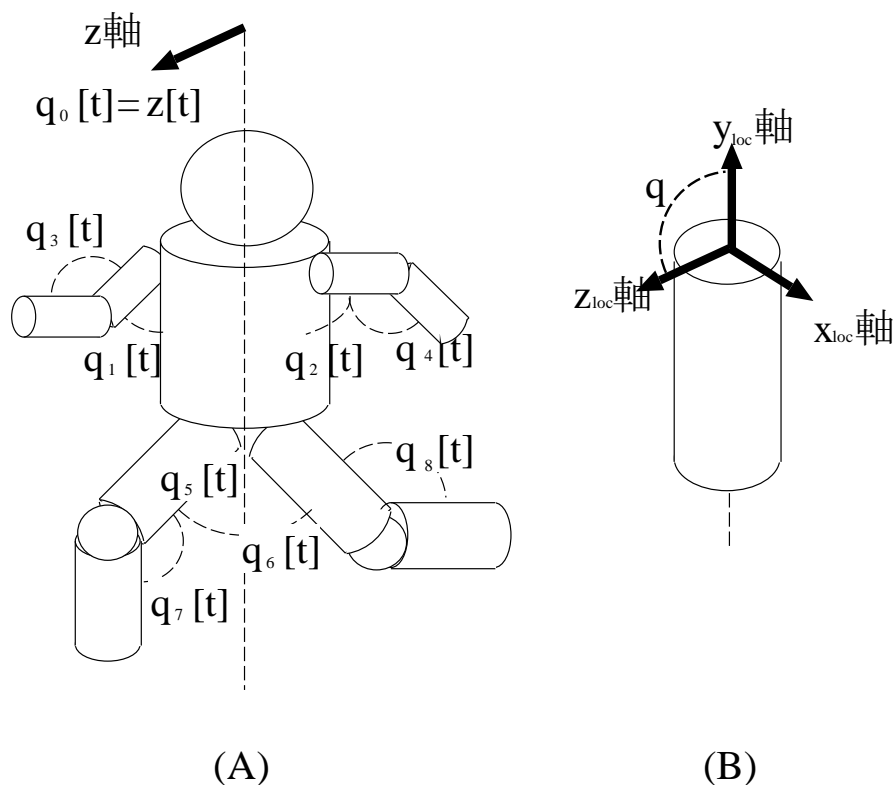


図 3.2 標準人体モデルとパラメータ

### 3.3 評価関数の設定によるモデルマッチングの定式化

人物領域から時系列の標準人体モデルのパラメータを得るために、3次元モデルマッチングを行なう。本研究では、時系列全体で人物領域と標準人体モデルの照合度を計算し、モデルパラメータを求める。各フレームで照合度を計算し、各フレームごとにパラメータを求める手法も考えられるが、あるフレームで腕などが胴体などによって完全に隠蔽された場合、肘などのパラメータを得ることができない。しかし、各パラメータの時系列の変化が滑らかであると仮定すると、時系列全体で評価しているため、前後の複数のフレームから補間することが可能になり、正しいパラメータを求めることが可能である。

この3次元モデルマッチングの評価関数を式(1)に示す。この評価関数を最小にするパラメータを解とする。第1項は、各パラメータによって作成される標準



人体モデルのアニメーションを2次元平面へ投影した時系列の領域と、時系列の人物領域との照合度を評価する項である。また、第2項はパラメータ全体の動きの滑らかさを評価する項である。

$$E = -W_m \sum_{t=0}^{N-1} E_{model}(\mathbf{p}_t) + W_s \sum_{i=0}^{P-1} E_{spline}(\mathbf{q}_i) \quad (1)$$

ここで  $W_m$ ,  $W_s$  は重み係数,  $N$  は動画像のフレーム数,  $P$  はパラメータ数を表す。また,  $\mathbf{q}_i$  はパラメータ  $q_i$  を時系列に並べたベクトル,  $\mathbf{p}_t$  はある時刻  $t$  のすべてのパラメータをベクトルで示したものとする。つまり, 以下のように定義する。

$$\begin{aligned} \mathbf{q}_0 &= [q_0[0], q_0[1], \dots, q_0[N-1]] \\ \mathbf{q}_1 &= [q_1[0], q_1[1], \dots, q_1[N-1]] \\ &\vdots \\ \mathbf{q}_8 &= [q_8[0], q_8[1], \dots, q_8[N-1]] \\ \mathbf{p}_t &= [q_0[t], q_1[t], \dots, q_7[t], q_8[t]] \end{aligned}$$

$E_{model}(\mathbf{p}_t)$  は,  $\mathbf{p}_t$  によって姿勢が決定される標準人体モデルを2次元平面に投影した領域  $M$  (モデルシルエット) と, 時刻  $t$  の人物領域  $S$  との照合度を示すものである。照合度は図 3.3 に示すように。

$$E_{model}(\mathbf{p}_t) = 2 \times (S \cap M) - \bar{S} \cap M - S \cap \bar{M}$$

と定義する。