

因子分解法によるイメージベースモデリングの検討

能勢容幸† 齋藤隆文‡

†東京農工大学大学院 工学研究科

‡東京農工大学大学院 生物システム応用科学研究科

概要

さまざまな状況でのウォークスルーを行う3次元モデルをイメージベースモデリングで作成するためには、多くの画像を必要とする。そのため、手動操作が必要な手法ではユーザへの負担が大きくなるという問題がある。そこで本稿では、動画から物体形状とカメラ運動を同時に復元する因子分解法による、簡易なイメージベースモデリング手法を提案する。提案手法は建築物内部を対象として、因子分解法により得られた物体形状を基にモデリングを行う。建築物内部ではさまざまな視点から撮影できず、因子分解法により得られる物体形状には誤差が含まれる。しかし、作成する3次元モデルを直方体に限定することで、高精度なモデリングを行うことができる。

Examination of Image-Based Modeling by Factorization Method

Hiroyuki NOSE† Takafumi SAITO‡

† Graduate School of Technology, Tokyo University of Agriculture and Technology

‡ Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology

abstract

Image-based modeling needs many pictures for walk-through. Therefore, the cost of the method which needs manual operation is high. In this paper, we propose simple image-based modeling by the factorization method recovers shape and motion from video sequence. This method performs indoor 3D modeling based on the shape by the factorization method. The shape includes some errors, because some pictures cannot be taken indoors from various views. However, we can perform 3D modeling in high accuracy by creating only box model.

1 はじめに

実写画像を用いることで容易に3次元形状を構築するイメージベースモデリング&レンダリングが盛んに研究されている。そのなかで、QuickTime VR[5]やTour Into the Picture(以下TIP)[6]などが提案され、手軽に3次元モデルを作成することができる。QuickTime VRは複数の視点の異なる実写画像から円柱投影されたパノラマ画像を作成することで仮想環境のウォークスルーを可能にする。TIPは1枚の画像上で消失点等を指定することで簡単な3次元構造を仮定し、3次元アニメーションを作成する。

これらのイメージベースモデリング&レンダリングで、さまざまな状況でのウォークスルーを行う3次元モデルを作成するためには、多くの画像を必要とする。従来の手法での入力には1枚または複数枚の静止画を使用しているが、ビデオカメラで撮影した動画を使用することで撮影コストを削減することができる。しかし、情報量の多い動画を入力すると、手動操作が必要な手法ではユーザへの負担が大きくなるという問題がある。そのため、動画からでも手軽に3次元モデルを作成できる手法が必要になる。

そこで本稿では、ビデオカメラで撮影した動画から3次元モデルを作成する簡易なイメ

ージベースモデリング手法を提案する．提案手法は建築物内部を対象として，因子分解法により得られた3次元情報を基にモデリングを行う．

提案手法は，動画から選択・追跡された特徴点の2次元座標から，それらの3次元座標とカメラ運動を同時に復元する因子分解法を利用する．因子分解法により得られた特徴点の3次元座標を基に，作成する3次元モデルは一般的な屋内を表現できる直方体に限定してモデリングを行う．

ウォークスルーを行うためには，作成した直方体モデルを組み合わせることで廊下や部屋を表現し3次元仮想環境を作成する．

2 因子分解法

動画からの対象物体の3次元形状推定は，コンピュータ・ビジョンの大きな研究課題の1つである．Tomasiらが提案した因子分解法[1]は動画から選択・追跡された特徴点の2次元座標を入力として，正射影モデルに基づいてカメラ運動と物体形状を同時に計算することにより解を安定させる方法である．

また，Poelmanらは線形性を保ちながら中心射影に対する近似精度がより高い，疑似中心射影モデルの因子分解法[3]を提案した．提案手法ではこの疑似中心射影モデルの因子分解法を用いる．

2.1 疑似中心射影モデル

疑似中心射影のカメラモデルを示す．世界座標系の原点を復元する物体の重心Cに採り，特徴点pの世界座標系での3次元座標を \mathbf{s}_p とする．フレームfのカメラ中心の座標を \mathbf{t}_f ，画像面の2次元ローカル座標系の基底ベクトルを $\mathbf{i}_f, \mathbf{j}_f$ ，カメラの光軸方向を $\mathbf{k}_f = \mathbf{i}_f \times \mathbf{j}_f$ とする．点pの画像面での座標を (u_{fp}, v_{fp}) ，カメラの焦点距離は $l=1$ とする．疑似中心射影モデルでは， (u_{fp}, v_{fp}) は次式で表される．

$$\begin{aligned} u_{fp} &= \mathbf{m}_f \cdot \mathbf{s}_p + x_f \\ v_{fp} &= \mathbf{n}_f \cdot \mathbf{s}_p + y_f \end{aligned} \quad (1)$$

$$(p = 1, \dots, P; f = 1, \dots, F)$$

ただし，

$$\begin{aligned} z_f &= (-\mathbf{t}_f) \cdot \mathbf{k}_f \\ x_f &= \frac{(-\mathbf{t}_f) \cdot \mathbf{i}_f}{z_f}, y_f = \frac{(-\mathbf{t}_f) \cdot \mathbf{j}_f}{z_f} \end{aligned} \quad (2)$$

$$\mathbf{m}_f = \frac{\mathbf{i}_f - x_f \mathbf{k}_f}{z_f}, \mathbf{n}_f = \frac{\mathbf{j}_f - y_f \mathbf{k}_f}{z_f}$$

ここで z_f は物体重心Cの奥行きパラメータ， (x_f, y_f) は物体重心Cの中心射影による画像面への投影となる．

2.2 疑似中心射影モデルの因子分解法

疑似中心射影モデルに基づく因子分解法について概説する．(1)式から，計測行列 \mathbf{W} を定義する．

$$\mathbf{W} = \begin{bmatrix} u_{11} & \cdots & u_{1P} \\ \vdots & u_{fp} & \vdots \\ u_{F1} & \cdots & u_{FP} \\ v_{11} & \cdots & v_{1P} \\ \vdots & v_{fp} & \vdots \\ v_{F1} & \cdots & v_{FP} \end{bmatrix} \quad (3)$$

次に，各フレームに対して，全特徴点のx座標値の平均値 x_f とy座標値の平均値 y_f を求める．

$$\frac{1}{P} \sum_{p=1}^P u_{fp} = x_f, \frac{1}{P} \sum_{p=1}^P v_{fp} = y_f \quad (4)$$

そして， \mathbf{W} の各要素から(5)式のように x_f, y_f を差し引いて行列 \mathbf{W}^* を作る．

$$\mathbf{W}^* = \mathbf{W} - \begin{bmatrix} x_1 \\ \vdots \\ x_f \\ y_1 \\ \vdots \\ y_f \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \quad (5)$$

この \mathbf{W}^* はランクが 3 となるため次式のように分解できる．

$$\mathbf{W}_{(2F \times P)}^* = \mathbf{R}_{(2F \times 3)} \mathbf{S}_{(3 \times P)} \quad (6)$$

$2F \times 3$ の行列 \mathbf{R} はカメラの姿勢ベクトル $(\mathbf{m}_f, \mathbf{n}_f)$ $3 \times P$ の行列 \mathbf{S} は特徴点の位置ベクトル \mathbf{s}_p を，それぞれ並べた行列となる．実際には計測行列にノイズが含まれるため，行列のランクが 3 とは限らない．この場合でも特異値分解を用いて，大きな 3 つの特異値だけを保つように分解すれば最適分解が得られる．

しかし，(6)式の分解は一意ではない．実際，任意の正則行列 \mathbf{Q} により，

$$\begin{aligned} \mathbf{W}^* &= \mathbf{R}\mathbf{S} = \mathbf{R}\mathbf{Q}\mathbf{Q}^{-1}\mathbf{S} \\ &= (\mathbf{R}\mathbf{Q})(\mathbf{Q}^{-1}\mathbf{S}) = \hat{\mathbf{R}}\hat{\mathbf{S}} \end{aligned} \quad (7)$$

のように，無数の解が存在する．そこで，次の計量拘束と呼ばれる(8)式から \mathbf{Q} を求める．

$$\begin{aligned} \frac{\|\mathbf{m}_f^T\|^2}{1+x_f^2} &= \frac{\|\mathbf{n}_f^T\|^2}{1+y_f^2} \left(= \frac{1}{z_f^2} \right) \\ \mathbf{m}_f^T \cdot \mathbf{n}_f &= \frac{x_f y_f}{2} \left(\frac{\|\mathbf{m}_f^T\|^2}{1+x_f^2} + \frac{\|\mathbf{n}_f^T\|^2}{1+y_f^2} \right) \quad (8) \\ \|\mathbf{m}_1^T\| &= 1 \end{aligned}$$

そして， \mathbf{Q} を使って，

$$\mathbf{R} = \hat{\mathbf{R}}\mathbf{Q}, \mathbf{S} = \mathbf{Q}^{-1}\hat{\mathbf{S}} \quad (9)$$

とおけば，

$$\mathbf{W}^* = \mathbf{R}\mathbf{S} \quad (10)$$

に一意に分解できる．

3 提案手法

提案手法の主要な処理とアルゴリズムについて説明する．

3.1 特徴点追跡

提案手法で用いる因子分解法への入力は，動画から選択・追跡された特徴点の 2 次元座標である．動画からの特徴点追跡を行うため

に，文献[2]を基に Tomasi らが KLT Feature Tracker を開発した．KLT Feature Tracker は動画を入力とし，自動で特徴点を選択・追跡できる．ただし，動画のフレーム間で差が大きいと失敗する可能性があり，画像間の特徴点对応には未だ決定的な手法はない．よって，本稿では手動で特徴点追跡を行う．

特徴点として選択・追跡する点は，直方体のモデルを作成するため，その頂点となる 8 点が最低限必要である．追跡するフレーム数は，因子分解法の原理から 3 フレーム以上必要である．

追跡する特徴点の数とフレーム数は，多いほど物体形状復元の精度を高めることができる．

3.2 物体形状の復元

入力動画から選択・追跡した特徴点の 3 次元座標は，第 2 節で説明した因子分解法を用いて得られる．しかし，建築物内部ではさまざまな視点から撮影できないため，因子分解法により得られる物体形状には誤差が含まれる．そこで，Christy・Horaud により提案された手法[4](以下 CH 手法)を用いる．CH 手法は中心射影モデルによる像から疑似中心射影モデルによる像を反復推定し，疑似中心射影モデルの因子分解法の復元精度を高める手法である．

3.3 モデリング

CH 手法により得られた特徴点の 3 次元座標を基に，作成する 3 次元モデルは一般的な屋内を表現できる直方体に限定してモデリングを行う．正確には直方体ではなく，左右の壁，正面奥の壁，天井，床に相当する 5 つの面からなるモデルを作成する．

まず，モデルの頂点となる 8 つの点を決定する．特徴点追跡で直方体の頂点となる点が追跡されていれば，特徴点の 3 次元座標において，重心から最も遠くなる 8 点が頂点となる．決定した 8 つの頂点から，モデルを作成する．

しかし，CH 手法を用いたとしても特徴点

の3次元座標は誤差を含んでいるため、正確な直方体を作成することはできない。見た目に十分なモデルを作成できたとしても、建築物内部の廊下や部屋を表現するために作成したモデルを結合していくと不自然な結果になる。よって、頂点の3次元座標を補正し、正確な直方体モデルを作成する。

3.4 テクスチャマッピング

作成したモデルにマッピングするテクスチャは、入力動画を用いて作成する。

はじめに、テクスチャ作成に使用するフレームを入力動画から1枚選ぶ。次に、テクスチャを抜き出す領域を決定する。直方体の頂点となる2次元座標が追跡できているから、それらの2次元座標から領域を決定しテクスチャを作成する。5つのテクスチャを作成し、3つの壁と天井、床に相当する面にマッピングする。

3.5 アルゴリズム

(1) 特徴点追跡

入力動画から直方体の頂点となる8点を3フレーム以上で選択・追跡し、因子分解法への入力となる特徴点の2次元座標を得る

(2) 物体形状の復元

CH手法により、特徴点の3次元座標を得る

(3) モデリング

直方体の頂点となる8点の3次元座標を基に、左右の壁、正面奥の壁、天井、床に相当する5つの面からなるモデルを作成する

(4) テクスチャマッピング

入力動画から、追跡した2次元座標を利用してテクスチャを作成し、モデルの各面にマッピングする

4 実験

4.1 実験データ

実験には特徴点座標が既知の作成したデータを用いた。実験1として、図1に示すような高さ4m×幅2m×奥行き3mの廊下を用いてモデリング精度を検証する。実験2として、図6に示すような高さ4m×幅2m×奥行き7mの直方体と高さ4m×幅2m×奥行き5mの直方体を組み合わせたL字型の廊下のモデリングを行う。

4.2 モデリング精度

実験に用いた画像は9枚であり、カメラを平行・直進移動させて撮影した。第1画像と第9画像を図2,図3に示す。廊下の重心から10m手前、高さ2mの位置で、幅2mを等分して平行移動し5枚撮影、その後高さ2m、幅1mの位置から1mずつ直進し4枚撮影した。CH手法への入力となるデータは、特徴点数8、フレーム数9の2次元座標となる。

CH手法により得られた特徴点の表示結果を図4に示す。提案手法により得られた特徴点のみと、テクスチャマッピングしたモデルの表示結果を図5に示す。特徴点の3次元座標の誤差を次式で評価した。

$$\frac{1}{P} \sum_{p=1}^P \frac{\|\mathbf{s}_p^{true} - \mathbf{s}_p^{estimated}\|}{\|\mathbf{s}_p^{true}\|} \times 100(\%)$$

ここで、 \mathbf{s}_p^{true} は真値、 $\mathbf{s}_p^{estimated}$ は推定値である。

CH手法で得られた特徴点の誤差は、モデルの奥の壁となる4つの点と、手前の4つの点とでは異なった。奥の4点は12.5%、手前の4点は24.1%、平均18.3%の誤差となった。手前の4点は物体重心から広がるように復元され、直方体となるように特徴点の3次元座標が正確に求められなかった。

提案手法で得られた特徴点の誤差は、4.75%となった。正確な直方体となるように補正しているため、モデルの奥と手前の点で誤差に差はない。このように、直方体に限定

することで、精度の高いモデルを作ることができた。

4.3 モデルの結合

図6のようなモデルの場合には、2つの直方体モデルを結合して作成する。実験では高さ4m×幅2m×奥行き7mのモデルと高さ4m×幅2m×奥行き5mのモデルを作成した。2つのモデルにおいて、実験1と同様に用いた画像は9枚であり、同じ撮影方法でカメラを平行移動させて5枚、直進移動させて4枚撮影した。

2つのモデルを作成し、結合した結果を図7に示す。図6と図7を比較すると、提案手法で作成したモデルの奥行きが長い。結合する前の2つのモデル両方において、高さ・幅ともに約1%の誤差となったが、奥行きの長さは図6のモデルの約2倍となった。

提案手法により正確な直方体モデルを作成できるため、モデルを結合させても自然な結果が得られた。しかし、実験では奥行きの長さが正確に復元されなかった。

5 考察

実験1では高い精度でモデルを作成することができた。しかし実験2では、奥行きが正確に復元されなかった。実験1と実験2では、復元する廊下のモデルの奥行きが異なるだけで、撮影方法などの条件は同じである。

そのため、奥行きの長い廊下などのモデルを作成する場合には、分割して作成した小さなモデルを結合していく必要がある。それに伴い、実験で行ったようなカメラを平行または直進移動させて行う撮影を、分割しただけ多く行う必要がある。

今後、作成するモデルのサイズと撮影方法の関係について検討する。

6 おわりに

本稿では、因子分解法を利用したイメージベースモデリングについて検討した。因子分

解法により得られる物体形状が誤差を含んでいたとしても、屋内のモデリングを対象としているため、作成する3次元モデルを直方体に限定することで精度の高いモデリングが行える。

また、作成したモデルを組み合わせることで建築物の廊下や部屋を表現することで、建築物内部のウォークスルーを行うことができる。

参考文献

- [1] C.Tomasi,T.Kanade,"Shape and motion from image streams under orthography: a factorization method", IJCV, vol.9,no.2,pp.137-154(1992)
- [2] B.D.Lucas,T.Kanade,"An Iterative Image Registration Technique with an Application to Stereo Vision", Proc.7th IJCAI, pp.674-679(1981)
- [3] C.J.Poelman,T.Kanade,"A paraperspective factorization method for shape and motion recovery", IEEE Trans.PAMI,vol.19,no.3,pp.206-218(1997)
- [4] S.Christy,R.Horaud,"Euclidean reconstruction: from paraperspective to perspective",Proc.4th ECCV,vol2, pp.129-140(1996)
- [5] S.E.Chen,"QuickTime VR - An Image-Based Approach to Virtual Environment Navigation", Proc.SIGGRAPH '95,pp.29-38 (1995)
- [6] Y.Horry,K.Anjyo,K.Arai,"Tour Into the Picture: Using a Spidery Mesh Interface to Make Animation from a Single Image", Proc.SIGGRAPH '97, pp.225-232(1997)

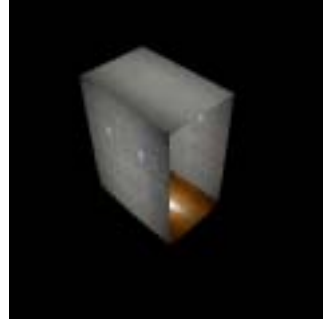
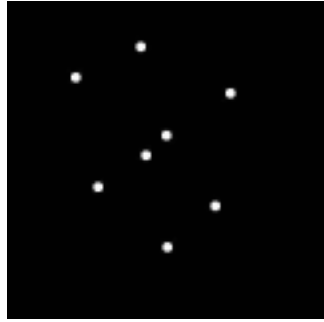


図 1.作成した特徴点のみのモデル(左)とテクスチャマッピングしたモデル(右)

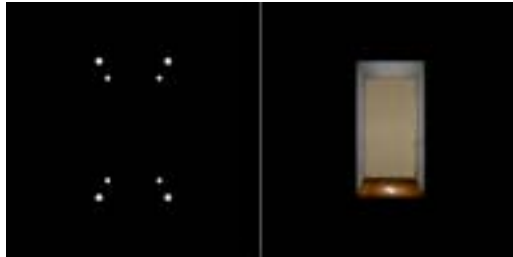


図 2.第 1 画像の特徴点(左)とモデル(右)



図 3.第 9 画像の特徴点(左)とモデル(右)

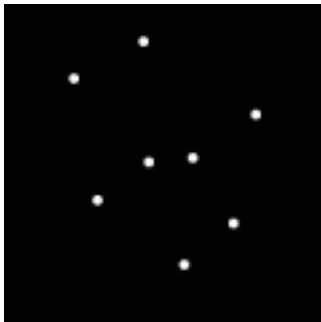


図 4.CH 手法により得た特徴点

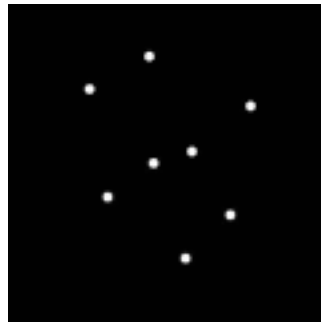


図 5.提案手法により得た特徴点(左)とモデル(右)



図 6.作成した L 字型廊下のモデル



図 7.提案手法で作成したモデル