

生成画像誤差最小化法 - CG Data from Motion

1 V - 1 1

森 偉明 齋藤 隆文 高橋 時市郎 大隅 規由 新谷 幹夫

NTT ヒューマンインタフェース研究所

1 まえがき

実画像をもとに、3次元CG用のモデリングデータを構築する手法を提案する。

従来の3次元情報入力手法は、両眼立体視法と運動立体視法の遮蔽による誤対応や情報の欠落 [2][1]、逆投影法のカメラパラメータ誤差と画像量子化誤差のため、3次元構造の抽出において、特徴点・領域の誤抽出と欠落が多く発生する [3]。これらの3次元構造としての誤りをいかに少なくするかという評価基準で、これまで多くの研究が行われてきた。

しかし、目的をCGに限れば、3次元構造に誤りがあっても、それを用いた生成画像が正しければ、差し支えない。そこで、提案手法では、カメラを並行移動しながら撮影した画像列を用いて、入力画像撮影時と同一のカメラパラメータを用いてCG技術により画像を再構成したときに、再構成画像と入力画像の差分が最小になるように3次元構造を抽出する。これにより、任意視点における再構成画像を、カメラ撮影を行なった現実世界に、見かけ上、最大限近づけることができ、高臨場感画像生成が可能となる。

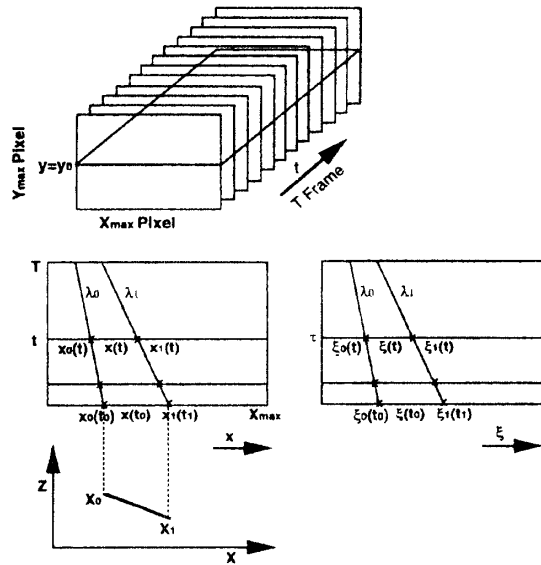


図 1: Principle

2 原理

今回提案する手法は、遮蔽が起こらないことを前提とする。

まず、3次元構造と生成画像の関係を明らかにし、入力画像列および抽出すべき3次元構造から誤差を計算する。図1に示す例を用いて誤差が計算可能であることを説明する。

図1において、入力時空間画像 $I(x, y, t)$ をスキャンライン方向 $y = y_0$ 面でスライスした画像（これをエビポーラ画像と呼ぶ）

$$f(x, t) = I(x, y_0, t) \tag{1}$$

がある。 x_0 と x_1 はエビポーラ画像 $f(x, t)$ 上の特徴点で、 λ_0 および λ_1 は特徴点 x_0, x_1 の軌跡である。カメラが、光軸に垂直かつ水平方向に等速直線移動したとすれば、特徴点の軌跡 λ_0, λ_1 は直線となる。この傾きおよびカメラパラメータから奥行きが計算できるので、エビポーラ画像上の特徴点を抽出し、さらに該特徴点の軌跡を推定することにより、該特徴点の軌跡の傾きより、該特徴点の奥行きが求まる [1]。

すなわち、これにより、像点 x_0, x_1 に対応する3次元線分 X_0, X_1 が抽出される。

次に、抽出された3次元構造 X_0, X_1 から画像を生成するプロセスを定式化する。まず、3次元構造から投影点を

*CG Data from Motion by Difference Minimization

Takeaki Y. Mori, Takafumi Saito, Toki Takahashi, Noriyoshi Osumi, Mikio Shinya

NTT Human Interface Labs.

求める処理を説明する。3次元線分 X_0-X_1 上の任意の点 X

$$X = \alpha X_0 + (1 - \alpha) X_1 \tag{2}$$

($0 < \alpha < 1$) の画像面への投影点 ξ は、カメラパラメータが与えられれば、既知の透視投影法を用いて計算できる。したがって、時刻 τ における入力画像の撮影カメラパラメータを用いれば、

$$\xi = P(X, \tau) \tag{3}$$

と表すことが可能である。また、定義から、両端点 X_0, X_1 の像点の軌跡 $\xi_0(\tau), \xi_1(\tau)$ は、特徴点の軌跡 λ_0, λ_1 と一致する。特に、 X の奥行き方向の変化が小さければ、線形補間

$$P(X, \tau) \simeq \alpha \xi_0(\tau) + (1 - \alpha) \xi_1(\tau) \tag{4}$$

により、簡略化することも可能である。

像点 ξ の色は、物点 X における面の傾き、反射率特性、光源の位置や属性、などから既知の画像生成手法から求めることができる。特に、反射特性が完全拡散的で、光源が変化しないとすれば、物体の色は視線方向に依らず一定となる。この場合は、既知のテクスチャマッピング手法を用いて、像点の色 f_{syn} は物体のテクスチャ $F(X)$ により

$$f_{syn}(\xi, \tau) = F(X) \tag{5}$$

$$\xi = P(X, \tau) \tag{6}$$

と計算できる。また逆に、テクスチャ $F(X)$ は適当な時刻 t_0 における入力画像をもとに、

$$F(X) = f(x, t_0) \tag{7}$$

$$x = P(X, t_0) \quad (8)$$

と定義できる。すなわち、特徴点の軌跡 λ_i および入力画像 $f(x, t)$ と生成画像 $f_{syn}(\xi, \tau)$ の関係が定式化された。特に、線形補間式 (4) を用いることができる場合には、

$$f_{syn}(\xi, \tau) = f(x', t_0) \quad (9)$$

$$\alpha = (\xi_1(\tau) - \xi) / (\xi_1(\tau) - \xi_0(\tau)) \quad (10)$$

$$x' = \alpha x_0(t_0) + (1 - \alpha)x_1(t_0) \quad (11)$$

と単純化できる。

生成画像 $f_{syn}(\xi, \tau)$ は特徴点軌跡 λ_0, λ_1 により与えられるので、生成画像と入力画像の絶対値誤差 h は

$$h(\lambda_0, \lambda_1) = \int_0^T \int_{x_0(t)}^{x_1(t)} |f_{syn}(x, t) - f(x, t)| dx dt \quad (12)$$

のように書くことができる。ただし、 $0 < t < T$ は入力画像の撮影時間である。また、 $x_0(t), x_1(t)$ はそれぞれ λ_0, λ_1 を表している。特に線形補間式 (4) が可能な場合は、

$$h(\lambda_0, \lambda_1) = \int_0^T \int_{x_0(t)}^{x_1(t)} |f(x', t_0) - f(x, t)| dx dt \quad (13)$$

$$x' = \alpha x_0(t_0) + (1 - \alpha)x_1(t_0) \quad (14)$$

$$\alpha = (x_1(t) - x) / (x_1(t) - x_0(t)) \quad (15)$$

とできる。

特徴点軌跡が n 本ある場合の誤差評価関数も同様に、 h の和

$$H(\lambda_0, \lambda_1, \dots, \lambda_n) = \sum_{i=1}^n h(\lambda_{i-1}, \lambda_i) \quad (16)$$

と特徴点軌跡 $\{\lambda_i\}$ から計算することが可能である。したがって、誤差評価関数 H を最小にする特徴点軌跡を探索することにより、再構成画像と入力画像の差分が最小になるように3次元構造を抽出できる。

3 アルゴリズム

カメラが光軸に垂直かつ水平方向に等速直線移動しながら撮影する。撮影された画像列がフレーム毎にメモリに蓄積され、図1に示す時空間画像が形成される。ここで、 X_{max} と Y_{max} は画像のサイズで、 T はフレーム数である。

まず、時空間画像をスキャンライン $y = y_0$ 面方向でスライスし、エピポーラ画像を作る。エピポーラ画像に対し、フィルタ処理で輝度変化の激しいところを特徴点として抽出する。エピポーラ画像上の特徴点軌跡は直線になる特性があるため、多数の特徴点もしくはその近傍を通過するような直線は特徴点軌跡である可能性が高いと言える。このような直線を、特徴点に対してハフ変換を施すことにより求める。得られた特徴点軌跡の候補直線をエピポーラ画像の左側より順番に l_1, l_2, \dots, l_m と表す。

次に、式16で示した誤差評価関数 H を最小にする N 個の特徴点軌跡、 $\{\lambda_i\}_{i=1,2,\dots,N}$ を候補直線 $\{l_j\}$ 探索する。この探索は動的計画法により効率良く実現できる。それを以下に説明する。

(1) 評価関数 $h_0(l_j)$ は l_j を最左端の特徴点軌跡 λ_0 としたときに該特徴点軌跡より左側の画素で発生する誤差であり、式 (13) を基に、 λ_0 の左のすべての画素を用いて計算する。すべての候補 l_1, \dots, l_m に対し h_0 を計算し、その結果を記憶する。

(2) i 番めの特徴点軌跡 λ_i を決定するのに必要な DP テーブルを作成する。評価関数としては、絶対値誤差式 (13) を基に定義する。

(3) 特徴点軌跡候補 l_e を最右端の特徴点軌跡とした場合に発生する、該軌跡より右側の画素に関する誤差の評価関数 h_e は同様に式 (13) を基に、 λ_e の右のすべての画素を用いて定義する。すべての候補 l_1, \dots, l_m について $h_e(l_j)$ を計算し、その結果を記憶しておく。

(4) 評価関数

$$H_N = h_0(\lambda_0) + \sum_{i=1}^{N-1} h_i(\lambda_i, \lambda_{i+1}) + h_e(\lambda_N) \quad (17)$$

を最小化する軌跡数 N を決定する。

(5) 同評価関数 (式 17) を最小化する $\lambda_i \in \{l_j\}$, $i = 1, 2, \dots, N$ を決定する。最後に既知の手法 [1] 求められた特徴点軌跡 $\lambda_1, \lambda_2, \dots, \lambda_N$ の傾きから各特徴点 x_1, x_2, \dots, x_N の3次元位置 X_1, X_2, \dots, X_N を求める。求めた3次元特徴点間のテクスチャについては、適当な時刻 t_0 における2次元特徴点 x_1, x_2, \dots, x_N 間の画素値を、式 (7) と (8) を用いてマッピングする。すべてのエピポーラ画像 ($y = 1, 2, \dots, Y_{max}$) に対して上述の処理を行なうと、カメラ撮影行なった世界の3次元構造が推定される。

4 まとめ

以上説明したように、カメラで撮影した画像より推定した3次元構造が、3次元構造を用いて、入力画像撮影時と同一のカメラパラメータで再構成した画像と入力画像との誤差が最小になるように推定される。これにより、任意視点における再構成画像を、カメラ撮影を行なった現実世界に、見かけ上、最大限近づけることができ、高臨場感画像生成が可能となる。今後は、本手法を遮蔽のあるシーンに拡張する予定である。

参考文献

- [1] R. C. Bolles, H. H. Baker, and D. H. Marimont. "Epipolar-plane image analysis: an approach to determining structure from motion". *IJCV*, 1(1):7-55, 1987.
- [2] T. Kanade, M. Okutomi, and T. Nakahara. "A multiple-baseline stereo method". In *Proc. DARPA Image Understanding Workshop*, pages 409-426, 1992.
- [3] T. Y. Mori, S. Suzuki, and T. Yasuno. "Structure and motion of 3D object from multi-views". *IEICE*, E78-D(12):1598-1606, 1995.