

## マルチビューによる高精度3次元構造抽出手法

2C-4

姚偉明 鈴木智 堀越力 安野貴之

NTT ヒューマンインタフェース研究所

## 1 まえがき

複数の画像を用いてシーン空間に対して3D Votingを行なうことによりシーン空間中に存在する物体の3次元構造を抽出する手法がある[1, 2]。これはまずシーン空間をボクセルに分割し、各ボクセル初期値をゼロにする。それからカメラレンズと画像上の特徴点を通る線(これからカメラ視線とよぶ)を用いて、カメラ視線の通るボクセルに対してある値を加算する(これからVotingとよぶ)。最終的に集積値の高いボクセルはカメラ視線が多く交差しているところであるので、そのボクセルを3次元特徴点とみなすことができる。しかしこの手法を実際のアプリケーションに適用するには2つの大きな問題がある。(1)カメラ視線を用いる3D Voting処理はカメラレンズ中心位置のずれと画像の量子化誤差に敏感である。(2)誤差の影響のため、正確な3次元特徴点を抽出するために多数枚の画像(約90枚、即ち90個ビュー以上)が必要である。

本論文はカメラレンズ中心位置のずれと画像の量子化誤差の3D Voting処理に対する影響のメカニズムを解析し、これらの誤差の影響は3D Votingを行なった後のボクセル空間に対して $\nabla^2 G$ フィルタリングを行なうことで改善でき、6個のビューで正確な3次元構造を抽出できることを理論と実験で示す。

## 2 3D Votingの誤差

まず画像上特徴点の量子化誤差のみを考える。

ワールド座標における画像上特徴点の量子化誤差 $\delta I = (\delta x_i, \delta y_i, \delta z_i)^T$ は式(1)で与えられる。

$$\delta I = \frac{\partial I}{\partial I_c} \cdot \delta I_c = R^T \cdot \delta I_c \quad (1)$$

ここで、 $\delta I_c$ はカメラ座標系における画像上特徴点の量子化誤差である。

図1(a)に示したように画像上の特徴点は $I$ から $I'$ にずれたとすると、 $L$ と $I'$ を通るカメラ視線上一点は $S$ から $S'$ にずれてしまう。これらの関係は式(2)に示される。

$$\frac{\delta x_i}{\delta x} = \frac{\delta y_i}{\delta y} = \frac{\delta z_i}{\delta z} = \frac{z'_i - z_i}{z' - z_i} \quad (2)$$

ここで、 $I'$ と $L$ は実際の画像上特徴点とカメラレンズ中心であるため、 $z'_i$ と $z_i$ は既知である。また $S'$ は実際の

のカメラ視線上の点であるため、 $z'$ も既知である。従って、カメラ視線のずれ $\delta S_a = (\delta x_a, \delta y_a, \delta z_a)^T$ は式(3)で与えられる。

$$\delta S_a = \alpha \cdot \delta I \quad (3)$$

ここで $\alpha = \frac{z' - z_i}{z'_i - z_i}$ は定数である。

次にカメラレンズ中心の位置ずれのみを考える。

ワールド座標におけるカメラレンズ中心の位置ずれ $\delta L = (\delta x_l, \delta y_l, \delta z_l)^T$ は式(4)で与えられる。

$$\delta L = \frac{\partial L}{\partial R^T} \cdot \delta R^T + \frac{\partial L}{\partial M} \cdot \delta M \quad (4)$$

ここで、 $\delta R^T$ と $\delta M$ はそれぞれカメラレンズ光軸の回転誤差マトリックスとカメラレンズ中心の並行移動誤差ベクトルである。

図1(b)に示したようにカメラレンズ中心は $L$ から $L'$ にずれたとすると、 $L'$ と画像上特徴点 $I$ を通るカメラ視線上の点は $S$ から $S'$ にずれてしまう。 $\delta S_a$ と同様に $\delta S_b = (\delta x_b, \delta y_b, \delta z_b)^T$ は式(5)で与えられる。

$$\delta S_b = \beta \cdot \delta L \quad (5)$$

ここで $\beta = \frac{z'' - z_i}{z_i - z'_i}$ は定数である。

## 3 誤差影響の低減

上述によりカメラ視線の総あたりのずれ $\delta S = (\delta x, \delta y, \delta z)$ は $\delta S_a + \delta S_b$ になる。もし $\delta S >$ ボクセルのサイズ $\delta V$ であれば、同一3次元特徴点を写った各画像上の特徴点を通るカメラ視線はこの3次元特徴点の存在するボクセル $V$ から外れ、隣接ボクセルがVotingされる。これは疑似3次元特徴点が抽出される原因となる。

ここで $P_a(x_a)$ 、 $P_b(x_b)$ と $P(x)$ をそれぞれ $\delta S_a$ 、 $\delta S_b$ と $\delta S$ のワールド座標系の $x$ 軸における分布とすると、 $P(x)$ は $P_a(x_a)$ と $P_b(x_b)$ の畳み込みになる。

$$P(x) = P_a(x_a) * P_b(x_b) = \int P_a(x-t) \cdot P_b(t) dt \quad (6)$$

$\delta I_c$ の分布は一様分布であるため[3]、式(1)と(3)から $P_a(x_a)$ は一様分布であることがわかる。ビューの数 $N$ が十分であれば式(4)から $\delta L$ の分布は正規分布になることがわかる。式(5)から $P_b(x_b)$ は正規分布であることがわかる。従って $P(x)$ は曲率の緩やかなカーブ形状になる。ところで、3D Votingを行なった後のボクセル空間のヒストグラムは $P(x)$ と重なるため、真の3次元特徴点は $(V - \delta S, V + \delta S)$ 範囲に存在することを仮定できる。そ

<sup>1</sup>Accurate 3D Structure from Multi-Views by Filtered 3D Voting

Weiming YAO, Satoshi SUZUKI, Tsutomu HORIKOSHI, Takayuki YASUNO  
NTT Human Interface Labs.

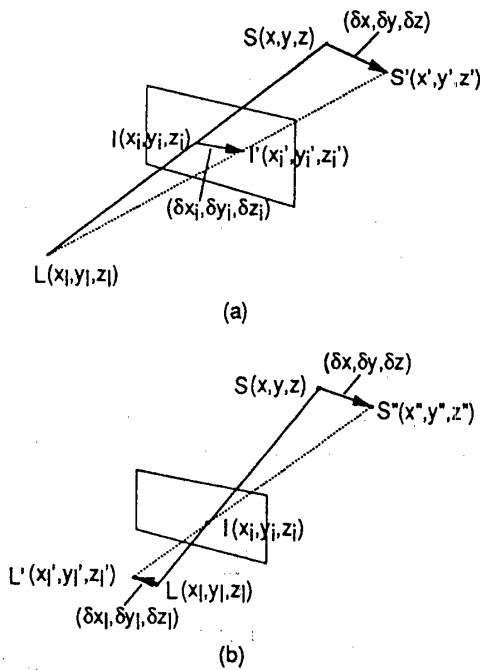


図 1: 誤差による 3D Voting への影響

ここで、ボクセル空間に対して  $2 \cdot \delta S$  のスケールの Laplacian フィルタをかけ、 $P(x)$  のピーク部分を鮮鋭化することで、容易に適切な閾値で真の 3次元特徴点の存在するらしきボクセルを抽出することができる。

実際に本研究は少ない画像数(即ち少ないビュー数)で 3次元構造を求めたいので  $\delta L$  の分布は離散的な分布に退化するため、単純に Laplacian フィルタをかけると離散的なノイズが抽出されてしまう。実際では式(7)に示した  $\nabla^2 G$  フィルタリングを行なうことで離散的なノイズを抑え、良好な 3次元構造を抽出することができた。

$$H' = H * G_1 - H * G_2 \quad (7)$$

ここで、 $H$  は 3D Voting を行なった後のボクセル空間のヒストグラムで、 $H'$  は  $\nabla^2 G$  フィルタリングを行なった後のボクセル空間のヒストグラムである。 $G_1$  と  $G_2$  は 2 つスケールの異なる Gaussian フィルタである。

#### 4 実験結果と考察

カメラを自由に動かしながらビデオレート 30 frames/second で 120 枚の画像を撮った。画像のサイズは  $360 \times 252$  ピクセルである。等間隔に 6 枚の画像を選び、Canny フィルタを用いて画像上の特徴点を抽出した。 $600 \times 600 \times 600\text{mm}$  のシーン空間を  $480 \times 480 \times 480$  個のボクセルに分割した。

図 2(b) は従来の 3D Voting 手法による 3次元特徴点抽出の結果である。この図からわかるように多数の疑似 3次元特徴点が抽出された。これは前述のように、カメラ視線のずれがボクセルサイズより大きい場合は、Voting score は真のカメラ視線が通るべきボクセルの周辺に分散され、真の 3次元特徴点が存在するらしきボクセルを抽出するための適切な閾値を得られないためである。

実際ではカメラ視線を用いてワールド座標の  $z$  軸に沿ってボクセル空間に対して 3D Voting を行なうため、 $\delta x$  と  $\delta y$  軸方向のカメラ視線のずれがボクセル空間のヒストグラムに大きい影響を与えるので、 $z$  軸に沿ってボクセル空間の各層のボクセルに対して 2次元フィルタリング  $\nabla^2 G(x, y)$  を行なった。その結果を図 2(c), (d) に示した。これにより、 $\nabla^2 G(x, y)$  フィルタリングにより、6 枚の画像(即ち 6 個のビュー)から正確な 3次元特徴点を抽出できることがわかる。

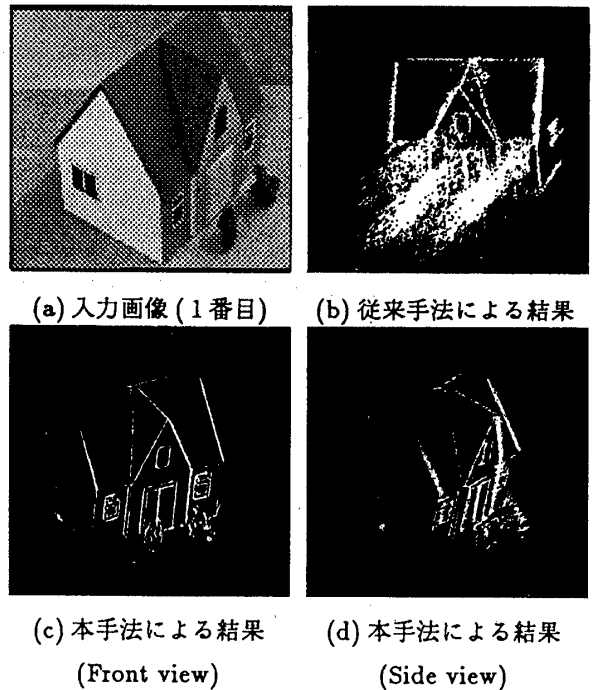


図 2: 実験結果

#### 参考文献

- [1] T. Hamano, T. Yasuno, and K. Ishii. "Direct estimation of structure from non-linear motion by voting algorithm Without tracking and matching". In *Proc. of ICPR'92*, volume 1, pp. 505-508, 1992.
- [2] W. Yao, T. Horikoshi, T. Yasuno, and S. Suzuki. "Structure from motion using coarse to fine 3D voting". In *Proc. of MVA'92*, pp. 313-316, 1992.
- [3] M. Morimoto, S. Akamatsu, and Y. Suenaga. "A high-resolution Hough transform using variable filter". In *Proc. ICPR'92*, volume 3, pp. 280-284, 1992.