

# 全方位視覚センサを用いた高解像度 3D モデリング

長原 一<sup>†</sup> 浜田 博 昭<sup>†</sup>  
八木 康 史<sup>†</sup> 谷内田 正彦<sup>†</sup>

ロボットの遠隔操作の遅延補償やウォークスルーといったロボティクスやヴァーチャルリアリティの応用のため、実環境を計算機に入力し仮想環境として再現することが注目されている。本研究では全方位視覚センサ HyperOmni Vision により移動しながら撮像された画像列から、環境の三次元幾何モデルおよびテクスチャを抽出することにより、三次元環境をモデル化する手法を提案する。本手法では、イメージモザイクングおよび画素ずらしによる超解像の考え方にに基づき、三次元幾何モデルの各面に対応するテクスチャ情報を高解像度化することを特徴とする。

## High-resolution Modeling of 3D Environment Using Omnidirectional Image Sensor

HAJIME NAGAHARA,<sup>†</sup> HIROAKI HAMADA,<sup>†</sup> YASUSHI YAGI<sup>†</sup>  
and MASAHIKO YACHIDA<sup>†</sup>

Recently, many applications in virtual reality and robotics, such as the compensation of time delays in tele-presence and walk-through, require to use virtual environment extracted from real scene. We propose a 3D environment modeling method which extracts a 3D geometrical model and texture images from an omnidirectional image sequence. However, the omnidirectional image sensor, HyperOmni Vision, has an intrinsic problem where the angular resolution of the sensor is lower than that of conventional video cameras. To solve this problem, we improve the resolution of its textures by using the technique of image mosaicing and super-resolution from the omnidirectional image sequence.

### 1. はじめに

ロボットの遠隔操作の遅延補償やウォークスルーといったロボティクスやヴァーチャルリアリティの応用において、実シーンからのリアルな仮想空間の構築は重要な技術課題で、様々な手法が提案されてきた。従来提案された手法の多くは、イメージペーストレンダリング (IBR) に基づく手法と、三次元環境モデルを用いるモデルペーストレンダリング (MBR) 手法に大別される。

IBR に基づく手法とは、実シーンをビデオストリームとして蓄積し、撮影時の視点位置での映像を再生する方法である。山口らや高橋らは、全方位視覚センサを入力手段とすることで、車などで走行しながら撮像した全方位ビデオストリームから任意視点での透視変換画像を表示することのできるシステムを構築している<sup>1),2)</sup>。この方法は、実画像を直接用いているため環

境モデルを必要としないという利点がある。しかし、撮像視点から大きく離れた場合、IBR により多視点画像列から生成した画像には、歪みが現れ、特に広視野画像の生成には問題が残る。さらに歪みを解消する方法として、空間に対しランダムな視点位置での映像を密に撮影する方法が考えられるが、膨大な画像を蓄積しなければならない。

これに対し、MBR による手法では、実画像を用いて三次元環境を復元または近似的にモデリングし、仮想視点画像を提示しようというアプローチがとられてきた<sup>3),4)</sup>。このアプローチは、任意の視点での広視野角映像が忠実に生成でき、さらに三次元モデルから仮想視点の画像を構築できるため、少ないデータで広域の再現が可能などの特徴を持つ。この方法においても、全方位視覚センサは一度に周囲 360 度のシーンがモデリングできることから利用されてきた<sup>5),6)</sup>。我々も、全方位入力画像列から環境地図およびロボット自己位置姿勢を同時推定する手法<sup>7)</sup>を提案している。しかし、全方位視覚センサは周囲 360 度が一度に撮像できる特徴により容易に三次元モデリングが行えるという利点

<sup>†</sup> 大阪大学大学院基礎工学研究科  
Graduate School of Engineering Science, Osaka  
University

と同時に、1 枚の CCD により全方位を撮像するため、入力画像の角度分解能が低いという欠点を持つ。そのため、得られた三次元幾何モデルにテクスチャマッピングしたとしても解像度が低く、利用者への提示目的には不十分といえる。

本研究では、全方位視覚センサ HyperOmni Vision<sup>8)</sup>により移動しながら撮像された画像列から、環境の三次元幾何モデルおよびテクスチャを抽出することで三次元環境をモデル化する手法を提案する。特に本手法では、イメージモザイクングおよび画素ずらしによる超解像の考え方より、モデリングに用いる画像列から観測点の関係を考慮することで、三次元幾何モデルの各面に対応するテクスチャの高解像度化を行う。本手法<sup>9)</sup>と類似した手法は、川崎ら<sup>6)</sup>によりほぼ同時に提案された。彼らの手法では、等速直線運動を仮定のもと EPI 解析により全方位画像列から三次元モデルを構築し、バックプロジェクションによりテクスチャの高解像度化を行っている。しかしながら、解像度特性が俯角に依存するといった全方位画像固有の性質は反映されていない。提案手法では、全方位視覚センサの解像度特性をあらかじめテーブル化した解像度重みテーブルを持つことで、異なる視点での画像列の位置合わせやバックプロジェクション<sup>10)</sup>に用いる画像の選択を行っている。

本論文とは異なる観点ではあるが、全方位視覚センサ自体の高解像度化のアプローチとして、我々はセンサの鉛直回転による画素ずれ画像列から内挿やバックプロジェクション、多重焦点化により高解像度化を行う手法<sup>11),12)</sup>を提案してきた。また、Nayer ら<sup>13)</sup>も横向きに設置したセンサを鉛直回転し、得られた全方位画像列から高解像度部位をモザイクングによりつなぎ合わせることで高解像度化を行う手法を提案している。しかし、これらの手法は固定視点による高解像度化であり、今回のような移動をととも画像列からの高解像度化とは異なる。

## 2. 前提条件

本研究では、モデリングを行う環境を、垂直平面に囲まれた移動物体を含まない静止空間とする。そして、この垂直平面をモデリングの対象とし、以後本論文では、対象平面と呼ぶ。モデリングおよびテクスチャ切り出しのためのランドマークは、環境中の垂直エッジとする。また、センサの動きを平面上の動きに限定し、照明の変化はないものと仮定した。

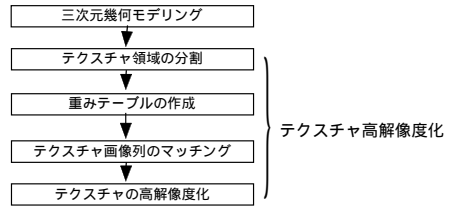


図 1 処理フロー  
Fig. 1 Processing flow.

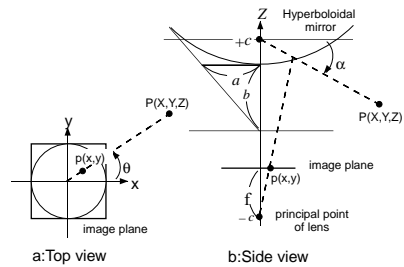


図 2 HyperOmni Vision の光学系  
Fig. 2 Optics of HyperOmni Vision.

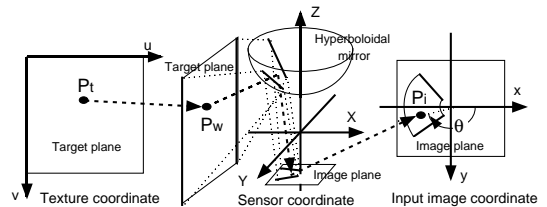


図 3 射影関係  
Fig. 3 Projective relation.

## 3. 処理概要

本手法の処理フローを図 1 に示す。個別の処理については 4 章でそれぞれ述べる。

本研究で用いる全方位視覚センサ HyperOmni Vision は図 2 に示すように、 $(X^2 + Y^2)/a^2 - Z^2/b^2 = -1 (Z > 0)$ ,  $c = \sqrt{a^2 + b^2}$  で表される鉛直下向きの双曲面鏡と鉛直上向きのカメラで構成される。ここで、 $X, Y, Z$  はセンサ座標を表し、 $a, b, c$  は双曲面固有のパラメータで、図 2 に示すように漸近線と焦点の関係を表す。この構造から周囲 360 度を同時観測することができる。環境中の任意の点  $P_w(X, Y, Z)$  は、図 3 および式 (2), (3) に示す関係で入力画像に撮像され、入力画像上の点  $P_i(x, y)$  と一意に対応する。また式 (1) の関係より、対象平面の垂直エッジは高さにかかわらず同一方位を示し、入力画像中では画像中心から放射上の線分として現れる。すなわち、入力画像中の放射エッジを検出することで、センサ全周の

すべての垂直エッジの方位を容易に推定できる．さらに，センサ移動にともなう垂直エッジの方位変化をトラッキングすることで，各垂直エッジおよびセンサ位置を推定することができる．ここで，推定された垂直エッジ間を対象平面とする．対象平面のテクスチャ座標  $(u, v)$  は，対象平面上の二次元座標を示し， $v$  軸は垂直エッジと平行， $u$  軸は  $v$  軸と直交な座標系とする．テクスチャ座標中の任意の点  $P_t(u, v)$  は式 (4) のように表され，センサ座標系  $P_w(X, Y, Z)$  に対応する．ここで， $[u_x, u_y, u_z]^T$ ， $[v_x, v_y, v_z]^T$  および  $[c_x, c_y, c_z]^T$  は，それぞれセンサ座標系での  $u$ ， $v$  ベクトルと交点を表し，エッジ端点位置より求められる．この関係より，入力画像中からテクスチャ画像を抽出することができる．次に，各対象平面はセンサ移動により様々な距離で撮像したことになる．そこで，異なる奥行きで観測された各面のテクスチャ情報から最適な画素を選択し，選択されたテクスチャに対しバックプロジェクション<sup>10)</sup>により個々の高解像度テクスチャを推定する．

$$\tan \theta = \frac{Y}{X} = \frac{y}{x} \quad (1)$$

$$x = Xf \frac{b^2 - c^2}{(b^2 + c^2)Z - 2bc\sqrt{X^2 + Y^2 + Z^2}} \quad (2)$$

$$y = Yf \frac{b^2 - c^2}{(b^2 + c^2)Z - 2bc\sqrt{X^2 + Y^2 + Z^2}} \quad (3)$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = u \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} + v \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} \quad (4)$$

## 4. 高解像度 3D モデリング

### 4.1 三次元幾何モデリング

図 4 に示すように基準座標系  $({}^0X, {}^0Y)$  に対するフレーム  $t$  のセンサ位置ならびに対象平面エッジ  $i$  の位置を  $({}^0X_t, {}^0Y_t)$ ， $({}^0X_i, {}^0Y_i)$  とする．ここで，基準座標系とは  $t = 0$  におけるセンサ位置とする．そして，フレーム  $t$  におけるセンサ座標系を  $({}^tX, {}^tY)$  とおき，センサ座標系に対する基準座標系の姿勢を  $\phi(t)$  とする．このとき，フレーム  $t$  におけるセンサ座標系から見た対象平面エッジ  $i$  の観測方位角  $\theta_i(t)$  には，式 (5) の関係が成り立つ．式 (5) より，時系列間で対象平面エッジ観測方位角  $\theta_i(t)$  を観測することで，各時刻の観測位置  $({}^0X_t, {}^0Y_t)$ ，姿勢  $\phi(t)$  ならびに  $i$  個の対象平面エッジ位置  $({}^0X_i, {}^0Y_i)$  が得られる．詳しくは参考文献 7) を参照されたい．ところで，式 (5) より対象平面のエッジ位置  $({}^0X_i, {}^0Y_i)$  とセンサ位置  $({}^0X_t, {}^0Y_t)$  が推定できれば，式 (6)，(7) よりフレ-

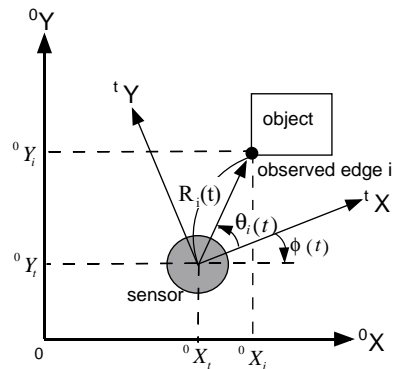


図 4 センサと対象平面エッジの位置関係  
Fig. 4 Relation between the sensor and the observed edge.

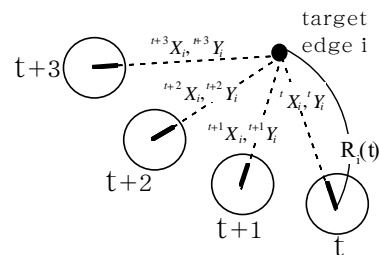


図 5 入力画像列と対象平面エッジの関係  
Fig. 5 Relation between input images and the observed edge.

ム  $t$  におけるセンサ座標系での対象エッジの相対位置  $({}^tX_i, {}^tY_i)$  が求まる．前章で述べたように，全方位視覚センサ HyperOmni Vision は透視投影の光学特徴を持つため，画像中のエッジ端点とセンサからの相対エッジ位置  $({}^tX_i, {}^tY_i)$  からエッジ端点の高さを求めることができる．しかし，単一の入力画像から正確に垂直エッジの端点を求められるとは限らない．そこで，図 5 に示すように，同一エッジが時系列間で観測できていることを利用し，各入力フレームでエッジ  $i$  を含む方位上の 1 ラインを切り出し，推定された位置に逆投影する．誤差なく 3D モデルとセンサ位置が推定できていれば，すべての視点から逆投影したエッジはすべて同じ高さとなり，両端は一直線になるはずである．しかし，実際には誤差を含むため，図 6 左に示すように，エッジの両端位置は一直線にならない．そこで，図 6 左の画像に対し， $Z$  軸方向に微分処理を施し (図 6 中央)，さらに  $t$  軸方向に投影することで，投影のピークを対象エッジの上下端点  $Z_i^H$ ， $Z_i^L$  として求める (図 6 右)．ところで，フレーム  $t$  における各エッジ位置  $({}^tX_i, {}^tY_i)$  は，地図生成結果  $({}^0X_i, {}^0Y_i)$  と各フレームのセンサ位置推定結果  $({}^0X_t, {}^0Y_t)$  から

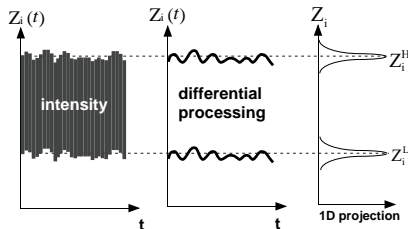


図 6 プロジェクションによる高さの決定  
Fig. 6 Edge height estimation.

計算することができる。しかし、両推定結果から計算したエッジ位置に各フレームで観測されたテクスチャを投影しても、必ずしもエッジのテクスチャがのってくるとは限らない。上述のエッジ端点の推定方法では、各フレームでのテクスチャがほぼ投影されることが重要となる。そこで本手法では、奥行きに関して地図生成結果  $({}^0X_i, {}^0Y_i)$  と各フレームのセンサ位置推定結果  $({}^0X_t, {}^0Y_t)$  から計算される相対距離  $R_i(t)$  (式 (8)) および、最も垂直エッジのテクスチャが反映された、エッジ抽出時の方位情報  $\theta_i(t)$  をテクスチャの抽出で利用した。

このようにして、すべての対象平面頂点座標  $({}^0X_i, {}^0Y_i, Z_i^H, Z_i^L)$  を求めることで、三次元幾何モデリングを行うことができる。

$$\tan(\theta_i(t) - \phi(t)) = \frac{{}^0Y_i - {}^0Y_t}{{}^0X_i - {}^0X_t} \quad (5)$$

$${}^tX_i = R_i(t) \cos(\theta_i(t)) \quad (6)$$

$${}^tY_i = R_i(t) \sin(\theta_i(t)) \quad (7)$$

$$R_i(t) = \sqrt{({}^0X_i - {}^0X_t)^2 + ({}^0Y_i - {}^0Y_t)^2} \quad (8)$$

#### 4.2 テクスチャの高解像度化

入力画像中のテクスチャの解像度は、撮像視点の位置によって異なる。また、HyperOmni Vision は放射状に環境を撮像するため、テクスチャは俯角に対しても大きな解像度差を持つ。そこで、これらの解像度差を考慮するため、画像列間のマッチングや高解像度化に用いる画素の選択指標として解像度重みテーブルを用いる。

##### 4.2.1 テクスチャ領域の分割

対象平面は、対象エッジ  $i, i+1$  を両端とする平面とする。このとき、4.1 節で求められたエッジ端点  $({}^tX_{i+1}, {}^tY_{i+1}, Z_{i+1}^H, Z_{i+1}^L), ({}^tX_i, {}^tY_i, Z_i^H, Z_i^L)$  ならびに式 (2), (3) から、入力画像に射影することでテクスチャ領域の切り出しを行う。モデルおよびセンサの推定位置は誤差を含むため、各フレームで切り出されたテクスチャ領域は、同じ位置および領域にはなっていない。そこで、次節で述べる解像度重みテーブル

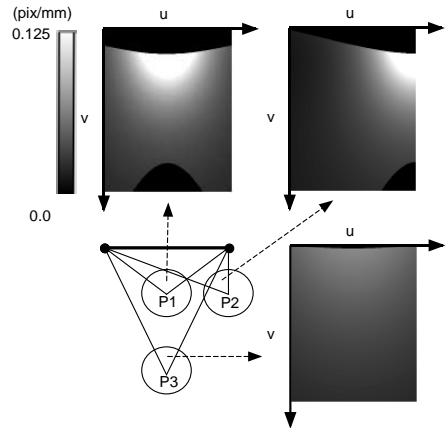


図 7 撮像位置に対する解像度重みテーブルの例  
Fig. 7 An example of resolution weighting table.

を用いることで、切り出されたテクスチャ領域間のより正確な位置合わせを行う。

##### 4.2.2 解像度重みテーブル

3 章で述べたように、対象平面上の任意の点  $P_t(u, v)$  と入力画像の点  $P_t(x, y)$  の関係は、式 (2) ~ (4) で表される (図 3)。したがって、この関係から、対象平面上での微小変動  $(\Delta u, \Delta v)$  が、入力画像の変動  $(\Delta x, \Delta y)$  ではどの程度となって現れるかを求めることができる。そこで、本手法では、様々な奥行きで撮影された入力画像の対象平面での解像度を、式 (9) に示す、対象平面上での微小変動の大きさに対する入力画像の変動の大きさとして定義し、解像度重みテーブルとして用いる。この関係は、センサと対象平面間の位置関係だけでなく、全方位視覚センサ HyperOmni Vision の射影特性をも反映している。図 7 は、P1, P2, P3 の 3 つの異なる視点から撮影した場合の解像度重みテーブルの例で、白い領域ほど解像度が高いことを示している。

$$W(u, v) = \frac{\sqrt{\Delta x^2 + \Delta y^2}}{\sqrt{\Delta u^2 + \Delta v^2}} \quad (9)$$

##### 4.2.3 テクスチャ画像列のマッチング

4.2.1 項の方法により切り出されたテクスチャ領域は、三次元モデルおよびセンサ位置に推定誤差が含まれるため、テクスチャ画像列間に微小な位置姿勢のずれが生じる。そこで、テクスチャを射影する対象平面上で各テクスチャ画像列を微小に動かしてテンプレートマッチングを行うことで、対象平面とセンサの相対位置を補正する。ところで、様々な位置から撮像された画像列から得られるテクスチャ画像は、部位により解像度に違いを持つ。そのため、解像度の違いを無視してマッチングを行うと、低解像度部分に引きずら

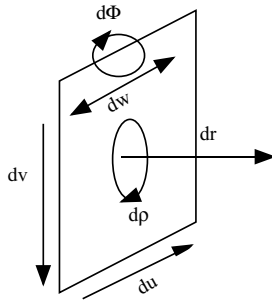


図8 テンプレートマッチングの変動パラメータ  
Fig. 8 Search region for template matching.

れ、高解像度部位のマッチング精度が悪くなる。また、4.2.4項の高解像度化においては、高解像度部位が優先的に用いられるため、高解像度部位を重視してマッチングを行う必要がある。そこで前述の解像度重みテーブル  $W(u, v)$  を用いて、式(10)の重み付き相関により濃淡値でマッチングを行う。補正のための微小変動パラメータは、図8に示すようにテクスチャ横方向  $dx$ 、縦方向  $dy$ 、センサからの距離  $dr$ 、幅  $dw$ 、Z軸回転  $d\Phi$ 、法線軸回転  $d\rho$  とした。

$$C = \sum_{u,v} W_{tr}(u, v) \times \frac{(f_t(u, v) - \bar{f}_t)(f_r(u, v) - \bar{f}_r)}{\sqrt{\sigma_{f_t} \sigma_{f_r}}} \quad (10)$$

$$\sigma_{f_t} = \sum_{u,v} W_{tr}(f_t(u, v) - \bar{f}_t)$$

$$\sigma_{f_r} = \sum_{u,v} W_{tr}(f_r(u, v) - \bar{f}_r)$$

$$\bar{f}_t = \frac{1}{N} \sum_{u,v} f_t(u, v), \bar{f}_r = \frac{1}{N} \sum_{u,v} f_r(u, v)$$

$$W_{tr}(u, v) = W_t(u, v) + W_r(u, v)$$

ここで、 $f_t(u, v)$ 、 $f_r(u, v)$  はフレーム  $t$  およびリファレンスのテクスチャ画像、 $W_t(u, v)$ 、 $W_r(u, v)$  はフレーム  $t$  およびリファレンスとなるフレームの解像度重みテーブル、 $N$  はテクスチャ画像の全画素数である。なお、初期リファレンスは、時系列画像のうち、対象平面に最接近したフレーム(最も解像度の良いフレーム)とし、以後、隣接フレームをリファレンスとして、テクスチャの位置合わせを行う。また、式(10)にて、リファレンスとなるテクスチャとあるフレームでのテクスチャ間でテンプレートマッチングを行うが、その際に用いる重みテーブル  $W_{tr}(u, v)$  は、各々の重みテーブルの和や積で表現できるが、今回は、予備実

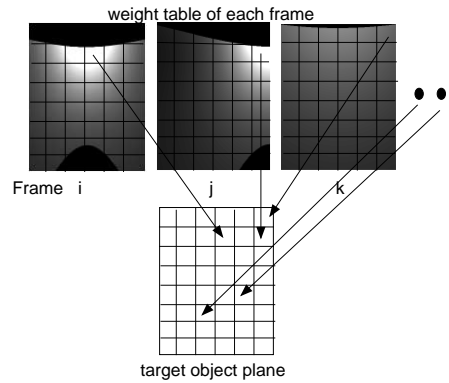


図9 重みテーブルによる入力フレームとフレーム数の選択  
Fig. 9 Selection of the frames number for super-resolution.

験から和による重み表現を選択した。

#### 4.2.4 テクスチャ画像の高解像度化

前述のように、対象平面からの距離により入力画像が持つテクスチャ解像度が異なる。画像列中で最大の解像度が得られるフレームは当然ながら図7のP1のように対象平面に近い視点から撮像されたフレームであるが、視野角が狭まるため上下が欠けてしまっている。そこで、モザイクキングの考え方により、各フレームの解像度の高い部位をつなぎ合わせて対象平面のテクスチャを生成する。この結果、視野角を保ちながら全体の解像度を向上させることができる。また、撮像された画像列は、センサ移動により画素ずれの効果が生じる。そこで本手法では、バックプロジェクションによる超解像手法<sup>10)</sup>によりさらなる高解像度化を行う。具体的に本手法では、図9に示すように、対象平面の各部位ごとに各フレームの解像度重みテーブルを比較し、解像度の高いフレームから順次選択する。そして、フレームの解像度重みの合計が閾値を超えるまで処理に用いることで、各部位ごとに最適なフレーム数を決定する。

バックプロジェクションによる超解像の手順を示す。初期テクスチャ画像  $f^0(P_t)$  から、式(11)より推定入力画像列  $g_t^0(P_i)$  を生成する。その推定入力画像列  $g_t^0(P_i)$  と実際に撮像された入力画像  $g_t(P_i)$  との差をとり、その誤差を関与した  $f^0(P_t)$  にバックプロジェクションし更新を繰り返すことで高解像度テクスチャを推定する(式(12))。ここで、 $P_t$ 、 $P_i$  はそれぞれテクスチャ画像および入力画像の画素である。 $f^n(P_t)$  は  $n$  回計算後の推定テクスチャ画像、 $g_k(P_i)$  と  $g_k^n(P_i)$  は実際の撮像画像とシミュレーション撮像画像である。 $P_{i_k, P_t}$  は、 $P_t$  に影響を与える  $P_i$  である。 $h^{PSF}$  はセンサシミュレーション<sup>11)</sup>により求めたPSF(Point

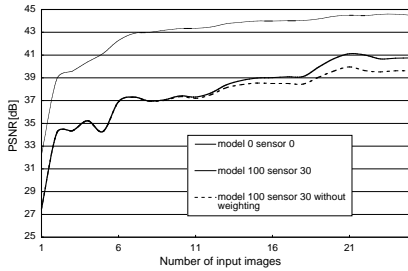


図 10 入力画像枚数に対する PSNR

Fig. 10 PSNR against the number of input images.

Spread Function) を二次元正規分布に近似して用いた。バックプロジェクションカーネル  $h^{BP}$  は、 $P_t$  の  $P_i$  に対する寄与率であり、式 (11) からポーティングにより求めた。

$$g_k^n(P_i) = \sum_{P_t} f^n(P_t) h^{PSF}(P_t) \quad (11)$$

$$f^{n+1}(P_t) = f^n(P_t) + \sum_{P_i \in \cup_k P_{i_k}, P_t} \frac{(g_k(P_i) - g_k^n(P_i))(h_{P_t P_i}^{BP})^2}{c \sum_{P_i' \in \cup_k P_{i_k}, P_i} h_{P_t P_i'}^{BP}} \quad (12)$$

## 5. シミュレーション

シミュレーションにより多視点入力画像を作成して解像度の向上を定量的に評価した。2m×2m の対象平面を仮定して、センサ視点位置は高さ 1.28 m、対象平面からの距離 0.5 m ~ 2.5 m、対象平面との中心からのずれ -1m ~ 1m の 2m×2m 領域中のランダムな視点とした。解像度の向上を評価する指標として結果画像と理想画像の PSNR (Peak Signal Noise Ratio) を用いた。図 10 にモデル誤差がない理想的な場合について高解像度化に用いる最大フレーム数に対する PSNR を示した。1 枚の入力画像から作成したテクスチャの PSNR 約 25 dB に対して、本手法は 40 dB 以上と解像度が向上していることが確認できた。また、推定誤差としてセンサおよび 3D モデルの位置に対し、標準偏差で (30 mm, 100 mm) の誤差を与えた場合を太線で示した。これより、誤差により解像度向上の限界はやや下がっているものの、1 枚の入力画像からのテクスチャに比べて、約 15 dB の向上が確認できる。なお、図 10、図 11 中とも、model, sensor の次の数字がモデルとセンサ位置に与えた誤差の標準偏差の大きさを表している。さらに、本手法では、対象平面の各部位ごとに使用するフレーム数を重みテーブルを用いて選択する方法をとっているが、その効果を合わせて図 10 に示す。図中点線は、重みテーブルを用い

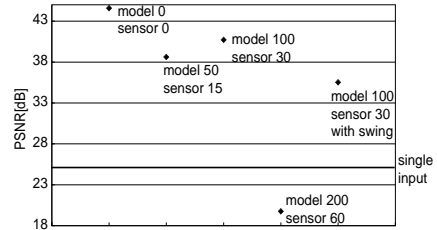


図 11 推定誤差と PSNR

Fig. 11 PSNR against standard deviation of location error.

かった場合の結果で、用いるフレーム数が少ない段階では差はないが、フレーム数が、10 フレームを超えるあたりから、提案手法に比べ、PSNR の上昇が鈍くなり、25 フレームの時点では約 1 dB の差がでた。このシミュレーション結果から、使用するフレーム数を重みテーブルにより制限することの効果を確認できる。

次に、三次元モデルの推定誤差の影響について、シミュレーションにより評価した。図 11 は、センサおよび 3D モデルの位置に対し、標準偏差で (15 mm, 50 mm), (30 mm, 100 mm), (60 mm, 200 mm) の位置誤差を与えた場合の PSNR の値である。図より、標準偏差 (15 mm, 50 mm), (30 mm, 100 mm) では、約 40 dB であったのに対し、(60 mm, 200 mm) の場合は、逆に約 20 dB と 1 枚から作成したテクスチャの 25 dB よりも悪化した。これは、誤差が大きすぎたため完全なマッチングが行われなかった結果である。このことから、(30 mm, 100 mm) 程度の誤差までは、高解像度化が行えることが確認できた。なお、実際のロボットを用いた実験より、推定誤差は視野変化の極端に小さい進行方向前方を除けば、(30 mm, 100 mm) 程度に収まっていた。また、実際のセンサ移動時には、床面の凹凸による揺れが想定される。これまでに撮った映像から揺れの大きさを評価してみると通常の走行では、屋内で 0.33 度の振動があった。そこで、3D モデルの位置に対し、上下方向も含め、(30 mm, 100 mm) の誤差を加えることで、実質 ±2 度程度の揺れの状態を再現してみた。その結果を図 11 に示す。この場合も、約 36 dB と解像度の改善が行えており、揺れに対しても十分適応可能であることが示された。また実用的には、センサ自体の運動に関しては、ジャイロセンサの情報を利用することも考えられる。通常市販のジャイロセンサを用いた場合、0.1 度程度で姿勢推定ができることから、ジャイロセンサとの併用により、大きな揺れに対しても適用可能であると考えられる。



表 1 ミラーパラメータ  
Table 1 Mirror parameters.

a	42.1 mm
b	42.7 mm
c	59.97 mm
f	3.9 mm



図 12 実験モデリング環境  
Fig. 12 Experimental environment.

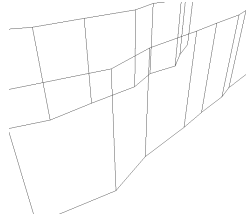


図 13 モデリング結果  
Fig. 13 A result of geometrical model.

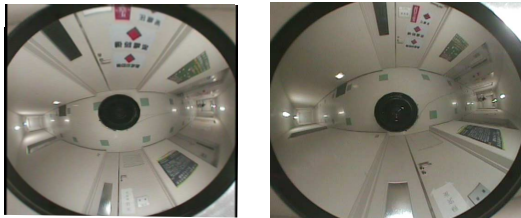


図 14 入力画像例  
Fig. 14 An example of input images. (left: near, right: far)

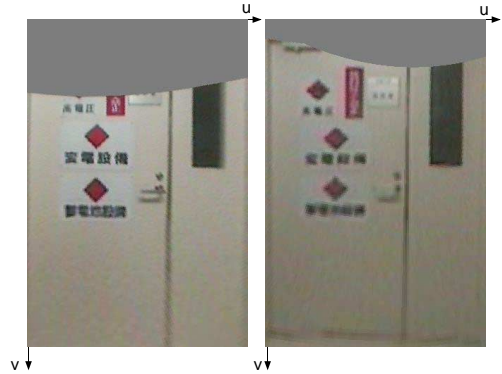


図 15 入力画像 1 枚からのテクスチャ画像例  
Fig. 15 Texture images transformed from a single image. (left: near, right: far) (350×404 pixels, 0.23 pixel/mm, 1 input image)

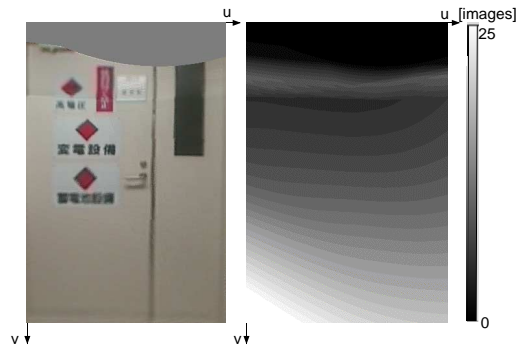


図 16 本手法によるテクスチャ画像例  
Fig. 16 A result of high-resolution texture and the number of images for resolution improvement. (350×404 pixels, 0.23 pixel/mm, 25 input images)

## 6. 実 験

移動ロボット (Nomad200/Nomadic Technologies Inc.) 上に表 1 のパラメータの双曲面鏡とビデオカメラ (EVI-310/Sony) で構成される全方位画像センサ HyperOmni Vision を搭載したシステムにより実験を行った。対象環境は図 12 に示すような、建物内の垂直平面の壁に囲まれた廊下とした。図 12 のようにロボットを廊下沿いに移動させながら撮像した 350 枚 (720 × 486 pixel) の多視点画像を入力画像列として用いた。得られた垂直エッジより手動で不要エッジの削除と対象平面を構成するエッジペアを選択してモデルを作成した。ここで、エッジはトラッキングにより関連付けられているため、全フレームにおいて削除および選択を行う必要はない。図 13 にモデリングの結果をワイヤフレームで示す。

次に 1 枚の対象平面を例示してテクスチャ高解像度化について考察する。図 14 に示すような対象平面に接近または離れた入力画像 1 枚からテクスチャを作成した結果を図 15 にそれぞれ示す。全方位視覚センサ



図 17 テクスチャ拡大画像  
Fig. 17 Zoom up result images. (left: proposed method, center: near, right: far)

HyperOmni Vision は垂直画角は制限されており、特に上方画角が狭い。そのため、大きな視野をモデリングするためには、図 14 右のように離れた視点で撮像することとなる。しかし、図 15 のように全体の解像度は当然ながら接近して撮像した図 14 左に対して低くなる。これらを含む 25 フレームを用いて本手法の高解像度化を行った結果を図 16 に示す。同図に高解像度化に使用されたフレーム数をテクスチャ部位ごとに示した。使用フレーム数は、すべてのフレーム中から解像度の高いものから順次選択し、選択されたフレー

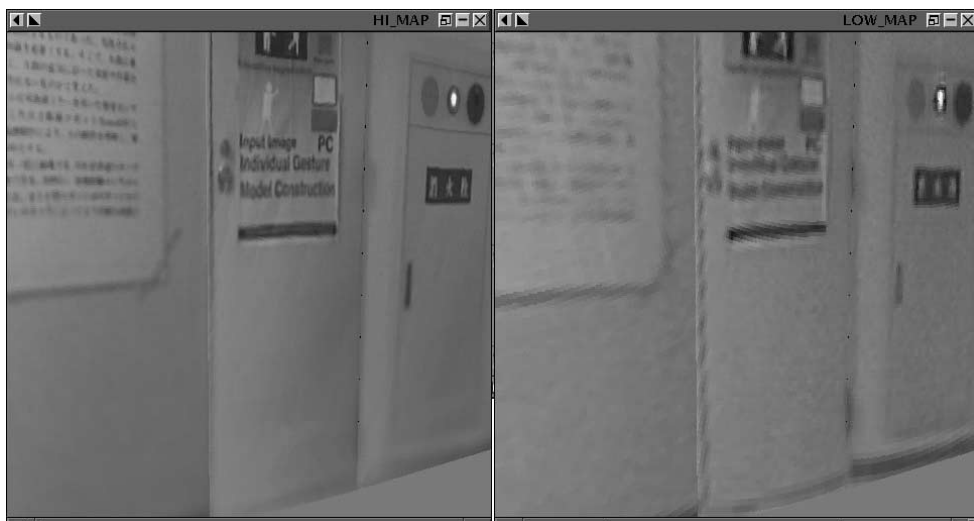


図 18 生成されたモデルを用いた仮想視点画像

Fig. 18 A comparative result of virtual view point image as built from generated 3D model.  
(left: high-resolution texture, right: single input texture)

表 2 計算コスト

Table 2 Computation cost.

Modeling	0.23	[sec/frame]
Matching	20	[min/frame/target plane]
Resolution improvement	6	[min/target plane]

Ultra Spark 450 MHz

ムの解像度重みの合計がある閾値を超えるまでのものとする。なお、使用フレーム数の最大値は 25 とした。また、解像度差を明確に示すため、図 15、図 16 の一部を拡大し、図 17 に示す。図 16 右のフレーム数から解像度重みテーブルにより画像下部や対象平面から離れて撮像される画像上部において、情報不足を補うように多くのフレームが高解像度化に用いられていることが確認できる。図 15 右は、図 16 の結果と同一視野範囲となるフレームであるが、明らかに図 16 の方が鮮明さが向上している。図 17 では、この結果がより顕著に分かり、右図では文字がまったく読めないのに対し、左図の本提案手法では「変電設備」と読むことができる。また中央の図との比較においても、菱形マークから直線がより直線らしくなっていることが確認できる。また図 15 左のフレームとの比較においては、本手法により広範囲のテクスチャが得られ、モザイクの効果が確認できた。さらに、表 2 に本実験での処理時間を示す。現状では特に、テクスチャの位置合わせのためのテンプレートマッチングに処理時間がかかっている。これは、今回 6 次元の探索空間を全探索により計算しているためで、今後は、位置合わせのための探索方法の改善をすすめる予定である。

これらの結果を用いて仮想視点からの映像を生成した例を図 18 に示す。同様に左図が本手法によりテクスチャを貼り付けたもの、右図がそれぞれ 1 枚の画像より生成したテクスチャを貼り付けたものである。この映像から仮想視点の映像においても解像度の向上が行えていることが確認できた。

## 7. まとめ

本論文では、全方位視覚センサ HyperOmni Vision により移動しながら撮像された画像列から、環境の三次元幾何モデルおよびテクスチャを抽出することにより、三次元環境をモデル化する手法を提案した。モデリングされた環境は、任意の視点での広角映像が忠実に生成でき、さらに三次元モデルから仮想視点の画像を構築できるため、少ないデータで広域の再現が可能などの特徴を持つ。また、本手法は、全方位ビデオストリームのみから環境モデリングが行えるため、撮像やモデル構築の時間と手間のコストを軽減できる。

さらに本手法では、イメージモザイクおよび画素ずらしによる超解像度化の考え方より環境モデリングに用いる画像列から高解像度化を行うことで、従来指摘されていた全方位視覚センサを用いた場合の低解像度問題を改善できた。今後は用いる特徴量を垂直エッジから一般的な特徴点に変更して、平面制約のないより複雑なシーンについて拡張したいと考える。また、処理の高速化もあわせて行いたいと考える。



## 参 考 文 献

- 1) 山口晃一郎, 山澤一成, 竹内治雄, 横矢直和: 全方位移動画像を用いた任意視点方向ステレオ画像の実時間生成と提示, 信学技報, PRMU-99-159 (1999).
- 2) 高橋拓二, 川崎 洋, 池内克史, 坂内正夫: 全方位画像によるレンダリングの手法, 情報処理学会研究報告, CVIM-199-5, pp.33-40 (1999).
- 3) Tomasi, C. and Kanade, T.: Shape and Motion from Image Streams under Orthography: A Factorization Method, *Int. J. Computer Vision*, Vol.9, No.2, pp.137-154 (1996).
- 4) 上原将文, 塩崎剛志, 全 炳東: 数値地図と動画画像解析による都市空間モデリング, 画像の認識・理解シンポジウム (MIRU200), Vol.2, pp.373-378 (2000).
- 5) 島村 潤, 山澤一誠, 竹村治雄, 横矢直和: 全周実画像とCGモデルの合成による仮想環境の構築, 画像の認識・理解シンポジウム (MIRU200), Vol.2, pp.367-372 (2000).
- 6) 川崎 洋, 池内克史, 池内正夫: 車載全方位カメラ映像からの建物画像の超解像度化, 情報処理学会研究報告, CVIM-125-13, pp.95-102 (2001).
- 7) 浜田博明, 八木康史, ネルスベンソン, 谷内田正彦: 全方位視覚センサを用いた環境地図とロボットの自己位置・姿勢推定法, システム制御情報学会論文誌, Vol.15, No.2 (掲載予定) (2002).
- 8) Yamazawa, K., Yagi, Y. and Yachida, M.: New real-time omnidirectional image sensor with hyperboloidal mirror, *Proc. 8th Scandinavian Conf. on Image Analysis*, Vol.2, pp.1381-1387 (1993).
- 9) 長原 一, 八木康史, 谷内田正彦: 全方位視覚センサを用いた高解像度3Dモデリング, 信学技報, PRMU-2000-152, pp.39-46 (2001).
- 10) Irani, M. and Pelg, S.: Improving resolution by image registration, *Graphical models and image processing*, Vol.53, No.3, pp.231-239 (1991).
- 11) 長原 一, 八木康史, 谷内田正彦: 画素ずれ全方位画像列を用いた高解像度化, システム制御情報学会論文誌, Vol.14, No.6, pp.322-329 (2001).
- 12) 長原 一, 八木康史, 谷内田正彦: 多重焦点全方位画像からの超解像度化, 日本ロボット学会学術研究予稿集, Vol.2, pp.783-784 (2000).
- 13) Nayer, S.K. and Karamarkar, A.: 360×360 Mosaics, *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (2000).

(平成13年3月1日受付)

(平成13年9月11日採録)

(担当編集委員 太田直哉)



長原 一

平成10年山口大学大学院理工学研究科博士前期課程修了。平成13年大阪大学大学院基礎工学研究科博士後期課程修了。現在、日本学術振興会研究員。画像処理、コンピュータビジョンの研究に従事。博士(工学)。



浜田 博昭

平成11年大阪大学基礎工学部卒業。平成13年同大学大学院基礎工学研究科修士課程修了。在学中、移動ロボットの研究を行う。同年4月(株)シャープ入社。現在、通信システム事業本部にて、ソフトウェア開発に従事。



八木 康史 (正会員)

昭和58年大阪大学基礎工学部制御工学科卒業。昭和60年同大学大学院修士課程修了。同年三菱電機(株)入社。同社産業システム研究所にてロボットビジョンの研究に従事。平成2年大阪大学基礎工学部情報工学科助手。同学部システム工学科講師を経て、現在、同大学大学院システム科学分野助教授。平成5年~6年オックスフォード大学客員研究員, IEEE, 電子情報通信学会, システム制御情報学会, 日本ロボット学会等各会員。工学博士。



谷内田正彦 (正会員)

昭和46年大阪大学大学院工学研究科修士課程修了。同年同大学基礎工学部制御工学科助手。助教授を経て同学部情報工学科教授, 平成6年同学部システム工学科教授。現在, 同大学大学院システム科学分野教授。昭和42年~43年デンマーク原子力研究所留学。昭和47年~48年米イリノイ大学にて Research Associate。昭和55年~56年西独ハンブルグ大学 Research Fellow。昭和57年米ミネソタ大学 CDC Professor。ロボット学会, 人工知能学会等会員。著書「ロボットビジョン」(昭晃堂, 大川出版賞受賞); 「コンピュータビジョン」(丸善, 編著)等。コンピュータ・ビジョン, 画像処理, 人工知能, 移動ロボット等の研究を行っている。工学博士。