

信頼性に基づく 3D 情報の重みづけ統合手法

森本正志 赤松茂 末永康仁

NTT ヒューマンインタフェース研究所

各初期視覚モジュールが出力する 3次元構造情報を 1つの仮説とみなし、その仮説に対する信頼性に基づいて行なわれる統合処理手法を提案する。ここでは特に、陰影系初期視覚モジュールの統合手法について報告する。ある点に対する複数の仮説の各々に対し、仮説の確からしさを信頼度として求める信頼度算出、信頼性の高い方を選択して統合処理を開始する仮説選択、信頼性の高い仮説から順に弛緩法処理を進めていく仮説伝播の統合処理機能の検討・実験を行なった。その結果、モジュール単独結果より安定した処理結果が得られることがわかった。

A Weighted-Integration Method of 3D Information using Certainty

Masashi Morimoto Shigeru Akamatsu Yasuhito Suenaga

NTT Human Interface Laboratories

1-2356 Take, Yokosuka, Kanagawa 238-03, Japan

There are various methods, called early visual modules, to extract 3D structure information from 2D images. Each module has a limited extraction capability according to the cue and algorithm it uses. This paper proposes a method for integrating these modules to make up for their limitations. The 3D structure information each module presents is regarded as a hypothesis, and the hypothesis integration is performed by referring to the certainty of the hypothesis. With this integration, more stable results are obtained than by using each module separately.

1 はじめに

使いやすいマンマシンインタフェースを提供するために視覚情報を使うことは有効である。我々は、人物の顔画像情報の認識・生成技術を用いてインタフェースの向上を図ることを考えている[1]。従来の顔画像認識手法においては顔の2次元的な特徴のみに着目することが多かったが、この特徴は顔の姿勢変化の影響を受けやすい。したがって、より高性能な認識を行なうためには3次元的な特徴にも着目することが必要であると考えられる[2]。この枠組における重要なファクターの1つに、2次元画像からの3次元復元処理をあげることができる。

この3次元復元のために必要となる処理は、おおきく初期視覚・中間視覚・高次視覚の3つのレベルに分けることができる[3],[4]。そのうち、3次元構造情報の特徴抽出を司る初期視覚機能[Early Vision]については、立体視・運動視・テクスチャ・輪郭・陰影などの手がかり(cue)に基づくさまざまな手法が研究されている[3],[5],[6]。しかし一般に、それぞれの単独の手法には適用範囲や抽出能力の限界が存在する。特に、顔画像のような複雑な形状物体の復元には、状況に応じた複数の手法の選択・制御・融合が必要であると考えられる。このような統合処理を司る中間視覚機能[Middle Vision]の研究は、現在進み出したところである。[4],[7],[8],[9],[10],[18]。

このような統合処理においては、通常2つ以上の手法が各々の手法の限界を補うように融合される。しかし、その各々の手法の出力情報は完全に正しい情報として扱われ、その確からしさについて考慮されていないことが多い。

本稿では、各初期視覚手法が出力する3次元構造情報を1つの仮説とみなし、その仮説に対する信頼性に基づいて行なわれる統合処理手法について検討を行なう。信頼性を基にした統合処理による高精度でロバストな復元の実現をそのねらいとしている。

以下、陰影系初期視覚手法及び統合処理のための機構について説明し、それらを用いた実験結果について述べる。

2 初期視覚モジュールの導入

本章では、陰影を用いる初期視覚手法の導入及び実験を行なう。人物の顔形状はおおむね滑らかであり、Shape from Shadingにおける滑らか拘束条件などを使いやすいこと、一方エッジのような特徴点を得にくいことが、陰影を用いる初期視覚手法を導入する理由である。ここでは、各手法のことをモジュールと呼び、各モジュールが出力する情報を仮説とみなす。画像上の各点に対し、法線ベクトルもしくは奥行き情報の仮説が生成される。

以下、3つの異なるモジュールの導入及び実験結果について述べる。

2.1 SSによる仮説生成

画像の陰影を用いて3次元復元を行なう手法としてよく知られているのは、Ikeuchi & HornによるShape from Shading(SS)法である[11]。この手法は、3次元物体の表面に対する滑らか拘束を仮定して、弛緩法により大局的に最も誤差の少ない状態を解とするものである(正則化)。SS法の概要は以下の通りである。

点 (x, y) における法線方向を表す関数を $f(x, y)$, $g(x, y)$ とする。近傍の法線ととの滑らかさを表す量 e_s を、

$$e_s = \int \int ((f_x^2 + f_y^2) + (g_x^2 + g_y^2)) dx dy \quad (1)$$

(f_x は x に関する f の1階偏微分、以下同様)、測定輝度値と理論値との誤差を表す量 e_i を、

$$e_i = \int \int (I(x, y) - R_s(f, g))^2 dx dy \quad (2)$$

($I(x, y)$ は輝度測定値、 $R_s(f, g)$ は反射率地図)とする。SS法は $e_s + \lambda e_i$ を最小化することで f と g を求めるとする手法である(λ : 重みづけパラメータ)。変

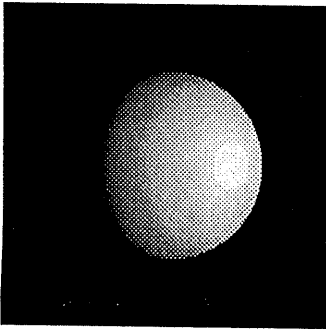
分法を用いて f, g を解いて離散化を行なうと、弛緩法として

$$\begin{aligned} f_{ki}^{n+1} &= \bar{f}_{ki}^n + \lambda(I_{ki} - R_s(f_{ki}^n, g_{ki}^n)) \frac{\partial R_s}{\partial f}, \\ g_{ki}^{n+1} &= \bar{g}_{ki}^n + \lambda(I_{ki} - R_s(f_{ki}^n, g_{ki}^n)) \frac{\partial R_s}{\partial g} \end{aligned} \quad (3)$$

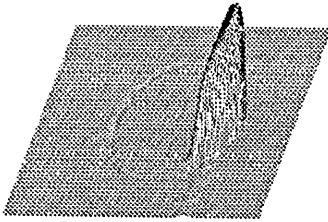
が得られる。 \bar{f} と \bar{g} は f と g の近傍の平均である。

初期条件としては、一般に遮蔽輪郭の法線方向情報が用いられる。

この SS モジュールの適用結果の例を図 1 に示す。この例では、右斜め光源下の球の CG 画像に対して行なった処理である。



原図 (右斜め光源下の球)



奥行き復元結果

図 1: SS モジュールの処理結果

図 1 の結果を見ると分かるように、右側の遮蔽輪郭周辺で仮説の生成が停滞している。このように、かなり単純な形状に対しても光源条件や弛緩法パラメータなどによって仮説の精度が悪化する。即ちこのモジュールの問題点として、仮説の精度や仮説生成の速度が対象形状・光源条件・弛緩法パラメータなどに依存することがあげられ

る。したがって、

1. 光源条件などに応じた仮説棄却の制御
2. 遮蔽輪郭とは異なった初期条件の導入
3. 弛緩法パラメータの制御

などが必要であると考えられる。

2.2 LSAM による仮説生成

繰り返し計算を行なう SS 法とは異なる一撃 (one-shot) 法として、Local Shading Analysis Method (以下、LSAM と略す) の導入を行なう。LSAM 法では、表面形状として球面を仮定して、画素の局所的な輝度変化情報に基づいて法線の方位角を推定する [12]。LSAM 法の概要は以下の通りである。

画像上での方向 ξ における微分値 d^2I は、 ξ だけ回転した座標系 (x^*, y^*) における x^* 方向の二次偏微分 $I_{x^*x^*}$ の値と等しい。この値は座標変換式より、 $I_{x^*x^*} = I_{xx} \cos^2 \xi + I_{yy} \sin^2 \xi + 2I_{xy} \cos \xi \sin \xi$ で求められる。 $I_{x^*x^*}$ が最小になる条件は、 $\frac{d(I_{x^*x^*})}{d\xi} = 0$ である。それより、

$$\tan 2\xi = \frac{2I_{xy}}{I_{xx} - I_{yy}}$$

となる。画像上の物体表面が球面であると仮定した際の I_{xx}, I_{xy}, I_{yy} の値を代入すると、

$$\tan 2\xi = \frac{2xy}{x^2 - y^2}$$

となる。一方、実際の球面上の法線の方位角 θ は、 $\tan \theta = y/x$ であるから、

$$\tan 2\theta = \frac{2 \tan \theta}{1 - \tan^2 \theta} = \frac{2xy}{x^2 - y^2}$$

である。これより $\tan 2\xi = \tan 2\theta$ が成立する。従って、局所表面形状を球面と仮定したとき、法線の方位角 θ は輝度の二次微分 d^2I の最小方向と等しい。

図 1 の原図に対する LSAM モジュールの適用結果を図 2 に示す。

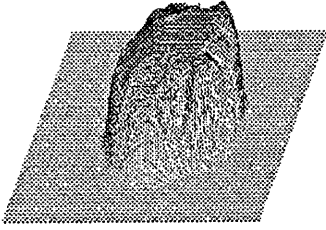


図 2: LSAM モジュールの処理結果

このモジュールの問題点として、方位角方向が一意に定まらないこと [14]、球面仮定やノイズに起因する仮説誤りが多いことがあげられる。したがって、光源条件などの入力条件に応じた仮説棄却の制御が必要であると考えられる。

2.3 LSS による仮説生成

別の一撃法である Linear Shape from Shading (以下、LSS と略す) を導入する [13]。この手法は、反射率地図を線形方程式で近似しようというものである。LSAM とは異なり局所的な微分情報を用いず、フーリエ変換を用いて奥行き情報を出力する。LSS 法の概要は以下の通りである。

光源方向を (τ, σ) と表し (球座標系)、物体の表面片の法線ベクトルを $(p, q, 1)$ と表したとすると、その表面片の (正規化された) 輝度は、

$$I(x, y) = \frac{p \cos \tau \sin \sigma + q \sin \tau \sin \sigma + \cos \sigma}{\sqrt{p^2 + q^2 + 1}}$$

で表される。この式を $p, q = 0$ のまわりで Taylor 展開すると、

$$I(x, y) \approx p \cos \tau \sin \sigma + q \sin \tau \sin \sigma - \frac{\cos \sigma}{2} (p^2 + q^2 - 2) \quad (4)$$

となる。2 次の項も無視することで、輝度式を法線ベクトルの線形式にすることができる。

ここで、 $I(x, y)$ のフーリエ変換式を $F_I(f, \theta) = m_I(f, \theta) e^{i\phi_I(f, \theta)}$ 、奥行き地図を表す関数 $z(x, y)$ のフーリエ変換式を $F_z(f, \theta)$ とすると、上記の近似

輝度式に基づけば

$$F_z(f, \theta) = \frac{m_I(f, \theta) e^{i(\phi_I(f, \theta) - \pi/2)}}{2\pi \sin \sigma f [\cos \theta \cos \tau + \sin \theta \sin \tau]}$$

と表すことができる。この式により、輝度情報から奥行き情報を復元することができる。

図 1 の原図に対する LSS モジュールの適用結果を図 3 に示す。

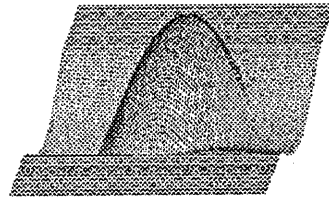


図 3: LSS モジュールの処理結果

このモジュールの問題点として、展開式の 2 次項以降が無視可能でなければモジュールが有効に働かない、ということがあげられる。すなわち、 $|p, q| \ll 1$ であり、光源方向が斜め方向であることが必要となる。したがって、光源条件などに応じたモジュール適用の制御、もしくは仮説棄却の制御が必要であると考えられる。

以上 3 つのモジュールの得失をまとめると、表 1 のようになる。

モジュール単独でさらに良い出力を必要とする場合には、ステレオ視 (複眼視) やフォトメトリックステレオ法 (複数光源) など、一般により多くの画像・拘束条件を用いるモジュールを導入する必要があると考えられる。

3 統合処理

2 章の各々のモジュールは、画像上の各点に対応する表面片の法線ベクトルもしくは奥行き仮説を生成している。この章では、それらの仮説の統合処理について述べる。前章での問題点に対処することで、統合処理により得られる結果が各モ

モジュール	SS	LSAM	LSS
手法	繰り返し法	一撃法(局所的)	一撃法(大局的)
出力	法線	法線	奥行き
適用域	滑らかな物体	局所的に球面	$ p, q \ll 1$, 斜めの光源
限界	遮蔽輪郭等のよい初期値が必要	微分値使用のためノイズに弱い	法線・光源条件が厳しい
問題点	入出力条件によるモジュール適用・パラメータ・仮説棄却などの制御が必要		

表 1: モジュールの得失

ジュール単独で得られる結果より改善されることを目標とする。

ここでは、各々の仮説に信頼性の大小が存在すると考え、それを基にした仮説統合の機構を提案する。まず、ある画素に対する複数の仮説の各々に対し、仮説の確からしさを信頼度として数値化する機構 [信頼度算出, 3.1] について述べる。次に、この複数の仮説から、信頼性の高い方を選択して統合処理を開始する機構 [仮説選択, 3.2] について述べる。最後に、信頼性の高い仮説から順に統合処理を進めていく機構 [仮説伝播, 3.3] について述べる。また、信頼度を含んだ最適化問題として統合処理を定式化する機構の検討も行なう。以下では、法線ベクトルという同じ出力形式を持つ SS と LSAM の統合を例に挙げて説明する。統合処理の概念図を図 4 に示す。

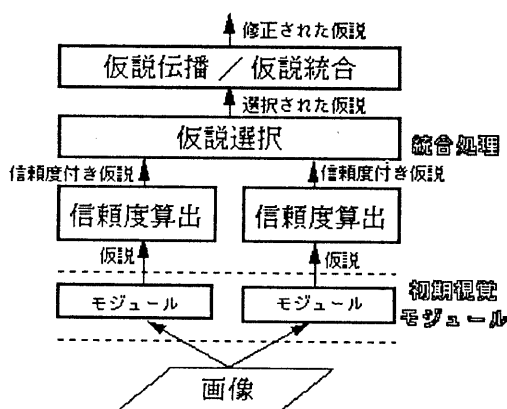


図 4: 統合処理概念図

3.1 信頼度算出

信頼度 (Certainty Factor) は、MYCIN [15] をはじめとするエキスパートシステムなどでよく取り入れられている手法の 1 つである。ここでは、各モジュールが出力する仮説が持つ確からしさ・信頼性 (reliability) の度合として信頼度を用いる。

信頼度の算出には様々な方法がある [8] が、基本的には各モジュールの持つ固有の情報を用いて算出する必要があると考えられる。算出方法として、

- モジュールが出力した仮説の値より信頼度を算出する方法
- モジュールの入力情報の誤差等より信頼度を算出する方法

並びにそれらの融合が考えられる。以下、これらについて述べる。

3.1.1 出力情報からの信頼度算出

まず、出力された仮説の値を用いて信頼度を算出する方法を検討した。例として、LSAM を挙げる。LSAM は局所処理であるため、近傍の仮説との整合性・類似度を信頼度の規準とした。LSAM の仮説に対する CF の算出式を

$$CF = \exp\left(\frac{f_x^2 + f_y^2 + g_x^2 + g_y^2}{-2\sigma^2}\right) \quad (5)$$

(σ は全画素にわたっての $\sqrt{f_x^2 + f_y^2 + g_x^2 + g_y^2}$ の平均値) で表す。概念図を図 5 に示す。

SS の遮蔽輪郭部分の仮説に対する CF も同様に算出できる。

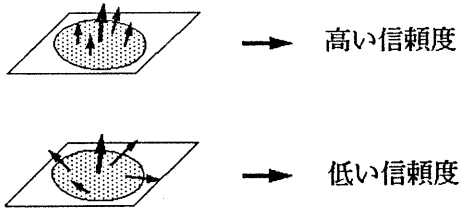


図 5: 仮説信頼度概念図

3.1.2 入力情報からの信頼度算出

前出の算出式は、近傍における法線方向の整合性に基いたものである。一方、法線や奥行き情報の算出過程そのものに対する信頼性の尺度として、Ikeuchi & Kanade の sensor detectability & reliability の考え方をを用いることができる [16]。LSS やフォトメトリックステレオなどの手法は入力条件や変換式がはっきりしているので、この方法による信頼度算出が容易であると考えられる。例として LSS を挙げる。LSS の精度が (4) 式の 2 次項の大きさに依存することから、例えば $CF = \sin \sigma$ とすることができる。また、

$$CF = \exp(-\cos \sigma(p^2 + q^2))$$

とすることで、光源条件と出力情報の両方を考慮することができる。

3.2 仮説選択

次に、仮説統合を行なうにあたり、ある点に対する複数の仮説のうちどれをどのように用いるかを司る仮説選択の機構について述べる。仮説選択により、各々のモジュールの欠点を補い合い、より正しい仮説を獲得することがこの機構のねらいである。

選択方法としては、

- 信頼性の最も高い仮説を用いる
- 単なる選択でなく、信頼性で重みづけをして新しい仮説を生成する

などが考えられる。

以下、SS と LSAM の 2 種類のモジュールを用い、1 つ目の方法を用いた場合について述べる。概念図を図 6 に示す。

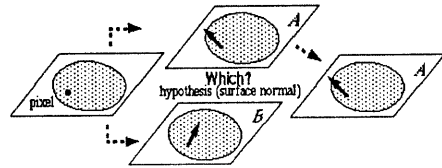


図 6: 仮説選択概念図

通常、画像の遮蔽輪郭部分は SS の仮説が、それ以外の部分は LSAM の仮説が選択される。これは、輝度変化が激しい遮蔽輪郭の部分では LSAM の仮説には誤差が大きくなると考えられるためである。

3.3 仮説伝播

次に、3.2 の仮説選択機構により生成された仮説を用いた仮説伝播の機構について述べる。この機構では、信頼性の高い仮説から順に弛緩法処理を進めていくことによって全体としての統合処理を行なう。これは、弛緩法の速度制御を行なうことに相当する。また、確からしい仮説を優先的に扱い、不確かな仮説部分は確かな部分と矛盾を起こさないように優先度を低くすることということに相当する。

以下、SS と LSAM の 2 種類のモジュールを用いた場合について述べる。弛緩法の式として、前出の (1)-(3) 式をそのまま用いることができる。この場合、3.2 と同様遮蔽輪郭部分における仮説が最も信頼性の高い仮説と考えられるため、その近傍から弛緩法の処理を開始する。それ以外の点では、近傍が弛緩法処理を始めるまでは処理を開始しないことで、弛緩法の速度制御を行なう。これにより、遮蔽輪郭部分の仮説が優先的に伝播されると考えられる。概念図を図 7 に示す。

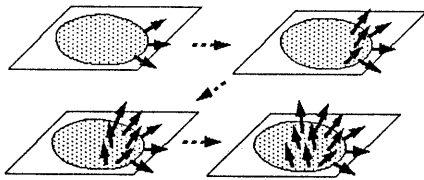


図 7: 仮説伝播概念図

3.4 仮説統合

3.3で述べたように、仮説の統合は弛緩法に基づいて行なう。弛緩法の式は、基本的に(1)–(3)式に従う。3.3の方法は、弛緩法の手速度制御を行なうことで統合処理に信頼度を組み込んでいると考えることができる。

一方、信頼度の値を直接弛緩法の式に組み込むことで、信頼度を含んだ最適化問題として統合処理を定式化することができる。また、LSSのように奥行き情報を出力するモジュールからの情報を組み込むためには、(1)–(3)式では不足である。そこで、以下のような定式化を考える^[17]。誤差量を、

$$\begin{aligned}
 E(z) = & \int \int z_{xx}^2 + 2z_{xy}^2 + z_{yy}^2 dx dy \\
 & + \Sigma \beta [z(x, y) - d(x, y)]^2 \\
 & + \Sigma \alpha_p [z_x(x, y) - p(x, y)]^2 \\
 & + \Sigma \alpha_q [z_y(x, y) - q(x, y)]^2 + \dots
 \end{aligned} \quad (6)$$

とし、 $E(z)$ を最小化するような弛緩法を行なう。ここで、第1項は表面形状の滑らか拘束、第2項以降はモジュールからの仮説(p, q, d)による拘束を表す。この(6)式を用いることで、奥行き・法線ベクトルの両方の仮説を基にした弛緩法処理ができる。また、 $\alpha_p, \alpha_q, \beta$ を仮説の信頼度で設定することによりその仮説の重みづけができる。

4 実験と考察

SSとLSAMの2種類のモジュールを用い、3.1.1, 3.2, 3.3で述べてきた統合処理機構を導入することで得られた結果を図8に示す。図1等で示したモジュール単独結果より安定した処理結果が得られ

ていることがわかる。ここでは(5)式による信頼度算出を行なっているが、この実験例の場合は妥当であったと言える。

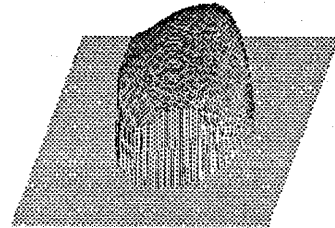


図 8: 統合処理結果

また、復元された奥行き情報と、図1の原図の理想的な奥行き情報との誤差2乗和と比較した結果を、表2に示す。

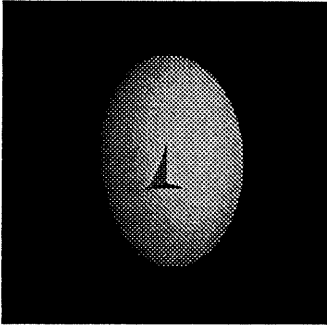
モジュール	SS	LSAM	LSS	統合
誤差2乗和	3.0	2.4	1.8	1.0

(統合時の誤差を1とする)

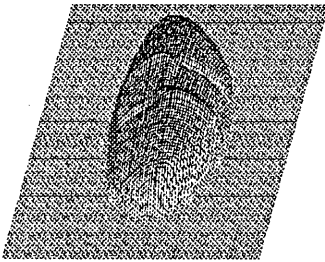
表 2: 復元誤差

次に、図9原図に対する処理結果を示す。図中の鼻の部分は、遮蔽輪郭とみなすことによる法線ベクトルも球面近似による法線ベクトルも正しい結果を与えることはできず、したがってSSやLSAMだけでは復元が困難である。そこで、 p, q が小さいと思われることや光源方向が斜めであることからさらにLSSを導入して、(6)式に基づいた統合手法で処理を行なった。図9の結果を見ると、ある程度まで形状は正しく復元されているが鼻の部分の形状が滑らかになり過ぎていることがわかる。この結果より、1つはLSSでは鼻の部分のような局所的な奥行き情報の変動を十分に復元できないこと、そしてもう1つとして(6)式における滑らか拘束が強過ぎるということが分かった。従って、ステレオ視などの局所的な奥行き情報の復元が可能なモジュールや、法線ベクトル・奥行き情報の不連続に対処した定式化が今後必要

であると考えられる。



原図 (CG 像、右斜め上の光源下)



奥行き復元結果

図 9: 実験結果

5 おわりに

各初期視覚モジュールが出力する3次元構造情報を1つの仮説とみなし、その仮説に対する信頼性に基づいて行なわれる統合処理手法について検討を行なった。この機構によれば、画像の内容に応じ他のモジュールからの情報の受け入れを可能とする拡張性がある。また、信頼性を用いることで高精度化・ロバスト化が期待できる。検討課題を以下にあげる。

- 信頼度の算出方法の定式化
- 最適化問題としての統合処理の定式化
- 仮説・信頼度の伝播方法の検討
- 複数の仮説・信頼度の競合 / 協調方法の検討

謝辞

本研究を進めるにあたり、日頃助言・議論いただく NTT ヒューマンインタフェース研究所視覚情

報研究部小林幸雄部長をはじめとする皆様に感謝致します。

参考文献

- [1] 赤松, 間瀬, 末永 “人物像を用いる優れたインタフェースの研究,” 信学春期全大, SA-7-1 (1990).
- [2] 赤松他, “3次元CGモデルを用いた顔画像のロバストな認識系実現への試み,” CV72-3 (1991).
- [3] Marr, D., 乾・安藤訳, “ビジョン: 視覚の計算理論と脳内表現,” 産業図書 (1987).
- [4] 乾敏郎, “視覚計算論における最近の諸問題,” MBE88-81, pp.1-10 (1988).
- [5] Aloimonos, “Visual Shape Computation,” Proc. of the IEEE, Vol.76, No.8 (1988).
- [6] Grimson, W.E.L., “From Images to Surfaces: A computational study of the human early visual system,” MIT Press (1981).
- [7] Poggio, T., Gamble, E.B. & Little, J.J., “Parallel integration of vision modules,” Science, 242, pp.436-440 (1988).
- [8] 松山隆司, “画像理解における推論方式,” 人工知能学会誌, Vol.4, No.1, pp.21-29 (1989).
- [9] 近藤栄治, 原島博, “輪郭情報と濃淡情報の統合による静止画からの構造抽出,” 信学秋季全大, A-151 (1990).
- [10] 大田友一, 渡辺正規, 角保志, “並列型画像理解-協調処理と並行トップダウン処理-,” 信学秋季全大, SD-9-1 (1989).
- [11] Ikeuchi, K. & B.K.P.Horn, “Numerical Shape from Shading and Occluding Boundaries,” Artificial Intelligence, Vol.17, Nos.1-3, pp.141-184 (1981).
- [12] Pentland, A., “Local shading analysis,” IEEE Trans. PAMI-6, No.2 (1984).
- [13] Pentland, A., “Shape Information From Shading: A Theory About Human Perception,” ICCV'88, pp.404-413 (1988).
- [14] 立部他, “光の方向に着目した Local Shading Analysis に関する一考察,” 信学技報 PRU89-122, pp.29-34 (1990).
- [15] Shortliffe, E.H., “Computer-Based Medical Consultation: MYCIN,” American Elsevier, New York (1976).
- [16] Katsushi Ikeuchi & Takeo Kanade, “Modeling Sensors: Toward Automatic Generation of Object Recognition Program,” CVGIP, 48, pp.50-79 (1989).
- [17] Demetri Terzopoulos, “The Computation of Visible-Surface Representations,” PAMI-10, No.4, pp.417-438 (1988).
- [18] 森本, 赤松, 末永 “3次元復元のための局所 / 大局情報の統合処理手法,” 信学秋期全大, SD-8-6 (1990).