

解説

3次元動画画像の理解

山本正信

電子技術総合研究所制御部視覚システム研究室 〒305 つくば市梅園 1-1-4

(1987年9月28日受理)

Dynamic Scene Analysis

Masanobu YAMAMOTO

Computer Vision Section, Electrotechnical Laboratory,
1-1-4, Umezono, Tsukuba 305

1. ま え が き

人間はその網膜に写った像の変化から、3次元世界の出来事を理解することができる。このような視覚の働きは人間のみならず多くの動物にとっても生存していくうえで欠くことのできない機能の一つである。3次元動画画像処理はこのような視覚機能を計算機上に実現することを目的としており、それは知能ロボットが一般的な環境下で行動を行なう際に必要となる技術の一つである。

動画画像処理の基本は連続する画像間の動きを把握することである。本文では、まず、画面上の動きを推定するいくつかの手法を紹介する。ついで、3次元動画画像処理の課題である、物体の3次元運動やシーンの3次元構造を推定する方法について述べる。最後に、知能移動ロボットへの応用例をあげておく。

2. 画像間の動きの推定

画面上の動きは、画像間の対応付けあるいは移動ベクトルによって表わすことができる。移動ベクトルは比較的小さな動きを表わす場合が多く、その画面上の分布はオプティカルフローと呼ばれている。

小さな動きの推定には勾配法が便利である。いま、画面上の点 (x, y) の移動ベクトル (v_x, v_y) は、画像の明るさの空間勾配 (E_x, E_y) と画像間の明るさの時間勾配 E_t を使って、近似的にはあるが次の簡単な方程式に拘束されることが知られている^{1,2)}。

$$E_x \cdot v_x + E_y \cdot v_y + E_t = 0 \quad (1)$$

この方程式を出発点として動きを定める方法が勾配法

である。 E_x, E_y, E_t は動画画像から差分演算などで容易に得られる。われわれが知りたいのは v_x, v_y であるが、(1)式だけからでは決定することはできない。そこで、異なる2点での移動ベクトルが等しいとすれば、それぞれの点で成り立つ上記の一次方程式を連立させることにより、 v_x と v_y を求めることができる。2点以上の多数の点で移動ベクトルが等しいとすれば、最小2乗法を使って、ノイズに強い結果が得られる。また、(1)式をパラメータ空間 (v_x, v_y) へHough変換したとき、2次元ヒストグラムのピーク点を探索することによっても移動ベクトルは得られる。この方法は複数個の動きの解析に便利である。

上記の条件を少しゆるめて、移動ベクトルは滑らかに分布すると仮定すれば、さらに複雑な移動ベクトル分布を求めることができる³⁾。すなわち、

$$\Sigma \{ (E_x \cdot v_x + E_y \cdot v_y + E_t)^2 + \alpha^2 (v_x^2 + v_y^2 + v_{yx}^2 + v_{xy}^2) \}$$

が最小となるように、移動ベクトル分布を求めるのである。ここで、上式の第1項は(1)式、 v_x と v_y の空間微分からなる第2項は滑らかさを表わす拘束に対応している。また、 α は両拘束間のバランスをとる重みである。この方法による移動ベクトルの決定例を図1に示す。

比較的大きな動きは、画像間のブロック同士を、相関法やデコンボリューション法、あるいは残差逐次検定法⁴⁾(SSDA)を使って照合することにより推定することができる。しかし、ブロック中の像が変形する場合には正確な対応付けがむずかしい。そこで、画像対象のモデ

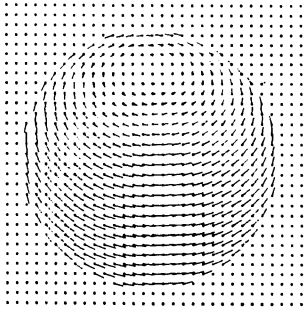


図1 推定されたオプティカルフロー³⁾

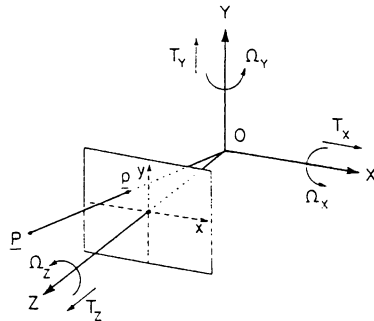


図2 3次元運動モデル⁹⁾

ルを使った照合法がオーソドックスな方法としてよく使われている⁵⁾。それは対象を領域の明るさや面積、周囲長などのような特徴量でモデル化し、画像間で特徴量を比較することにより対応付けを行なうのである。また、コーナー点などのような画像の特徴点を照合の単位として用いることも多く、画像間で対応付け可能な多くの候補のなかから、画像やシーンの性質を利用しつつ、弛緩法を使って一つに絞り込むことが行なわれている⁶⁾。

多数枚の動画画像を用いる方法として、画素の時系列間の相互相関解析を利用する方法⁷⁾や運動軌跡の画像化手法⁸⁾がある。前者は速度ベクトルの大きさを精密に測定することができる。また後者は運動物体の追跡の際に問題となるオクルージョンに強く、後で述べるように視点の移動による立体視にも使われている。

3. 3次元運動の推定

動画画像から物体の3次元運動を推定する場合、剛体の3次元運動パラメータを推定することが多い。それは、多くの動きがきわめて短い時間内には剛体運動とみなせること、長時間の動きも瞬間的な動きの積み重ねであることなどからうなずかれる。3次元運動パラメータを推定する方法は、前章で述べた画像間の対応付けを基礎とする2段階推定法と、対応付けを必要としない直接的推定法に大別される。

3.1 2段階推定法

物体面を平面と見なしたとき、面の模様によって引き起こされるオプティカルフローから面の傾きと3次元運動パラメータを求める問題を考えてみよう。この問題は、任意の物体面がいくつかの小さな平面領域で近似的にモデル化できることから、3次元動画画像解析の基本問題として古くから研究されてきた⁹⁻¹¹⁾。オプティカルフローと諸パラメータとの関係が非線形となるため、いかにしてこの非線形関係から解析解を得るかが永年の課題であったが、最近になって解が代数的に得られることが

示された¹²⁾。その方法を紹介しよう。

3次元座標空間をO-XYZとする。図2に示されるように、剛体の短い時間内での動きは、回転運動パラメータと並進運動パラメータによって表わされる。回転運動を $\Omega=(\Omega_x, \Omega_y, \Omega_z)$ 、並進運動を $T=(T_x, T_y, T_z)$ とする。そのとき剛体上の点(X, Y, Z)の移動ベクトル (V_x, V_y, V_z) は、

$$\begin{vmatrix} V_x \\ V_y \\ V_z \end{vmatrix} = \begin{vmatrix} 0 & -\Omega_z & \Omega_y \\ \Omega_z & 0 & -\Omega_x \\ -\Omega_y & \Omega_x & 0 \end{vmatrix} \begin{vmatrix} X \\ Y \\ Z \end{vmatrix} + \begin{vmatrix} T_x \\ T_y \\ T_z \end{vmatrix} \quad (2)$$

となる。

この剛体がXY平面に平行な面に原点を中心にして中心投影されるとする。簡単のため、焦点距離を1とおく。投影面の座標系をo-xyとおけば、剛体上の点(X, Y, Z)は(x, y)に次式で投影される。

$$(x, y) = (X/Z, Y/Z) \quad (3)$$

その点の移動ベクトル (v_x, v_y) は

$$\begin{aligned} v_x &= V_x/Z - X V_z/Z^2 \\ v_y &= V_y/Z - Y V_z/Z^2 \end{aligned} \quad (4)$$

で表わされる。(2)式を(4)式に代入して整理すれば、

$$v_x = -\Omega_x xy + \Omega_y(1+x^2) - \Omega_z y + (T_x - T_z x)/Z \quad (5a)$$

$$v_y = -\Omega_x(1+y^2) + \Omega_y xy + \Omega_z x + (T_y - T_z y)/Z \quad (5b)$$

を得る。

さて、剛体を平面と仮定しよう。平面の方程式は面の適当な1点までの距離 Z_0 と面の法線ベクトル (k_x, k_y) を使って、

$$Z = Z_0 + k_x X + k_y Y \quad (6)$$

と表わすことができる。このZを(5)式に代入し、さらに(3)式を使って整理すれば次式を得る。

$$v_x = a + bx + cy + dx^2 + d'xy \quad (7a)$$

$$v_y = a' + b'y + c'x + d'y^2 + d'xy \quad (7b)$$

ただし,

$$\begin{aligned} a &= T_x + \Omega_y, \quad b = k_x T_x - T_z \\ c &= -\Omega_z - k_y T_x, \quad d = \Omega_y + k_x T_z \\ a' &= T_y - \Omega_x, \quad b' = \Omega_z - k_x T_y \\ c' &= -T_x - k_y T_y, \quad d' = k_y T_x - \Omega_x \end{aligned}$$

ここで,

$$T_x = T_x/Z_0, \quad T_y = T_y/Z_0, \quad T_z = T_z/Z_0$$

さらに, $a \sim d'$ を使って, パラメータ $O_1 \sim O_8$ を次のように定義しておく.

$$O_1 = a = T_x + \Omega_y \quad (8a)$$

$$O_2 = a' = T_y - \Omega_x \quad (8b)$$

$$O_3 = b = k_x T_x - T_z \quad (8c)$$

$$O_4 = b' = \Omega_z - k_x T_y \quad (8d)$$

$$O_5 = (c + c')/2 = -(T_y k_x + T_x k_y)/2 \quad (8e)$$

$$O_6 = (c' - c)/2 = \Omega_z - (T_y k_x - T_x k_y)/2 \quad (8f)$$

$$O_7 = 2d = 2(\Omega_y + T_x k_x) \quad (8g)$$

$$O_8 = d' = k_y T_x - \Omega_x \quad (8h)$$

さて, 少なくとも4点での移動ベクトルが与えられれば, (7)式を連立させた線形方程式から8個の未知数 $a \sim d'$ を決定することができる. さらに, (8)式 of 非線形連立方程式を解けば, 3次元運動パラメータと面の傾きを知ることができる. $O_1 \sim O_8$ が与えられたとき, (8)式の未知数を逐次消去してゆけば, 結局 T_x に関する次の3次方程式を得る.

$$T_x^3 + C_1 T_x^2 + C_2 T_x + C_3 = 0 \quad (9)$$

ただし,

$$C_1 = O_3 + O_4$$

$$C_2 = -(O_5^2 - O_3 O_4 + A^2/4 + B^2/4)$$

$$C_3 = O_5 A B/2 - O_4 A/4 - O_3 B^2/4$$

$$A = O_1 - O_7/2, \quad B = O_2 - O_8$$

この3次方程式の根はすべて実根であり, それらを大きさの順に並べたとき真中の根が真の解であることが証明されている. (9)式の導出過程を逆にたどれば残りのパラメータが得られる.

3.2 直接的推定法

勾配法を使えば, 画像間の対応付けを介することなく, 動画像から直接運動パラメータを推定することができる. たとえば, 2.章で紹介した勾配法の最初の例は並進運動のパラメータを直接推定したものだといえる. 3次元運動はシーンの3次元形状についてのなんらかの情報がなければ, その動きを知ることができない. 対象を平面と仮定すれば, その見かけの動きは八つのパラメータを使って表現できる. このことは(7)式で見てきた. したがって, (7)式を勾配法の拘束式(1)に代入し

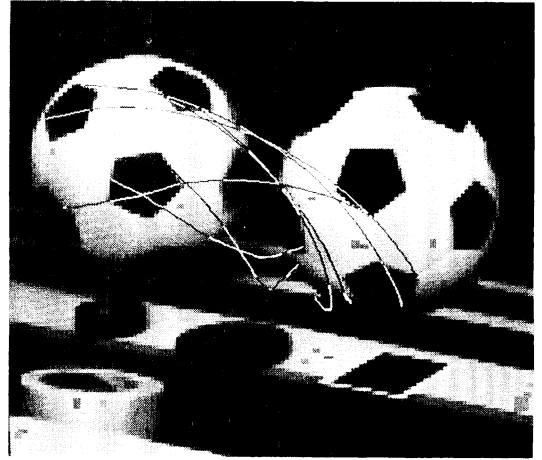


図3 3次元物体の追跡¹³⁾

て得られる, 次の線形推定式を介せば面の傾きと運動パラメータは同時に推定可能である.

$$\begin{aligned} E_x a + E_y a' + x E_x b + y E_y b' + y E_x c + x E_y c' \\ + x(x E_x + y E_y) d + y(x E_x + y E_y) d' + E_t = 0 \end{aligned} \quad (10)$$

すなわち, 物体面上の点に対応する八つ以上の画面上の点から導かれる線形連立方程式から, もしその連立方程式が独立ならば, 一意な解 $a \sim d'$ が得られる. さらに(8)式 of 非線形連立方程式を解くことにより3次元運動パラメータと面の傾きを知ることができる^{10,11)}.

さらに, 物体の形状情報が両眼立体視や幾何モデルなどにより与えられるとすれば, 線形連立方程式を解くだけで3次元運動パラメータが得られる¹³⁾. 簡単のために物体は xy 平面に正射影されるとする. そのとき, (2)式で表わされる (V_x, V_y) がそのまま画面上での移動ベクトル (v_x, v_y) となる. (2)式を(1)式に代入して整理すれば,

$$\begin{aligned} E_x T_x + E_y T_y - Z E_y \Omega_x + Z E_x \Omega_y \\ + (x E_y - y E_x) \Omega_z + E_t = 0 \end{aligned} \quad (11)$$

を得る. 物体までの距離 Z が与えられているので, (11)式は運動パラメータを未知数とする線形方程式となる. したがって, 同じ剛体上のいくつかの点から線形方程式を連立させることにより, その解を得ることができる. 実際に動画像から運動パラメータを推定しオプティカルフローを求めることにより物体を追跡した例を図3に示す.

4. 3次元構造の推定

物体のみならずカメラも移動しているとするれば, 前章での議論は物体とカメラとの相対的な3次元運動を求め

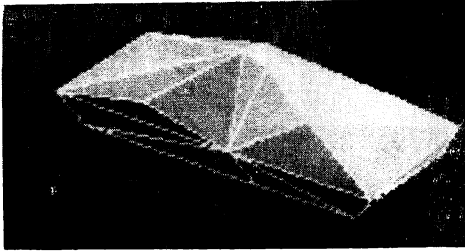


図 4 自動車の3次元形状¹⁴⁾

ていることになる。2段階推定法では3次元運動の推定時にその副産物として3次元形状を知ることができた。しかし多くの場合解を得る手続きが複雑であることや、解がノイズに敏感すぎるなどから、この理論を使って実際の動画像から3次元形状を復元した例はきわめて少ない。実際の動画像に適用するためには、未知数の数を少なくする意味で、相対運動の自由度を低減させることが望ましい。たとえば、図4に示す自動車の形状の復元例では、自動車の動きが道路面に拘束されていることが利用されている¹⁴⁾。相対運動が完全に既知の場合は、3次元情報を得ることのみが課題となり、この場合問題は伝統的なステレオ視の問題と一致する。

ステレオ視では左右の目で物体を見たときの視差がわかれば三角測量の原理で物体の3次元形状を知ることができる。問題は左右画像間の対応付けであるが、対応先はエッジ線上に拘束されるため、この線上を探索すればよく、一般の動画像解析における対応点探索ほどのむずかしさはない。対応の単位としてはエッジなどの特徴点やエッジではさまれた区間を用いることが多い。疎密解析法や弛緩法¹⁵⁾あるいは動的計画法¹⁶⁾などは、周囲の対応のコンテキストを考慮しつつ、数多くの対応付け可能な組合せの中から、最も良い対応付けを効率的に探索してくれる。しかし、両眼視だけでは対応付けの曖昧さは避けることはできない。また、一方の視点から見

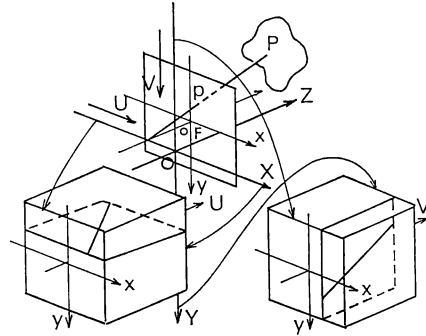
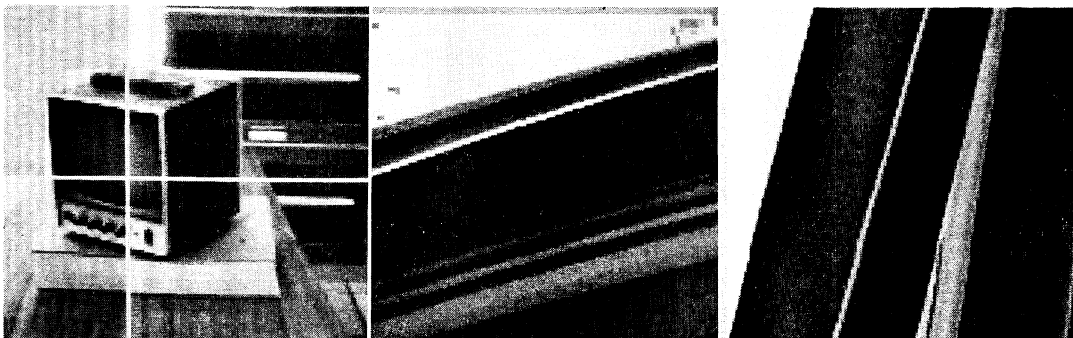


図 5 移動視点による3次元情報の抽出⁸⁾

ていた点が他方から見えなくなるというオクルージョンが起きている場合には対応付けがむずかしくなる。

3点以上の多数の視点から得られる多視点ステレオ画像を使えば、対応付けのあいまいさを解消することができる一方、オクルージョンによる対応付け不可能な点の検出も可能になる¹⁷⁾。さらに、視点を移動しつつ連続的に得られた動画像では、多視点ステレオ画像の長所を保存したまま対応付け処理の高速化と簡便化が可能になる。その手法⁸⁾を紹介しよう。

図5のように、シーンの座標系をXYZとする。シーンはXY平面に平行な投影面に中心投影される。いま、投影面の中心をX軸上に移動させながら、連続的に撮影した画像を堆積し3次元画像を構成する。そのとき物体上の点Pの像は、この3次元画像中の一枚の水平断面画像上に直線軌跡を描く。この直線の傾きが物体までの距離に比例する。したがって、直線の傾きを測定すれば、物体の3次元情報が復元できる。像が水平方向のエッジを有する場合には、測定結果の信頼性は低くなる。そこで、垂直方向にも投影面を移動させて、同様な3次元画像を構成し、その垂直断面画像から距離情報を得る。像のエッジの向きが垂直に近いときには水平断面画

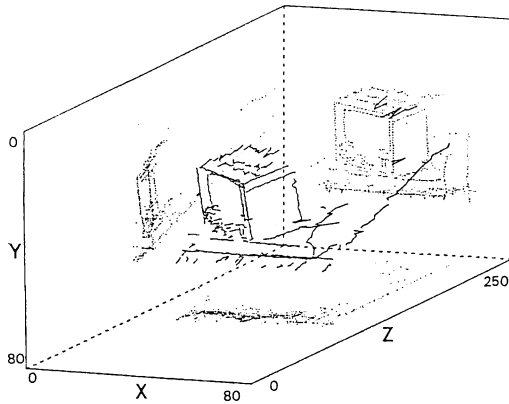
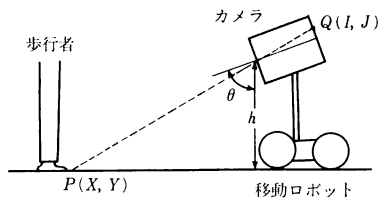


(a) 物体の画像

(b) 垂直断面画像

(c) 水平断面画像

図 6 物体と画像化された運動追跡⁸⁾

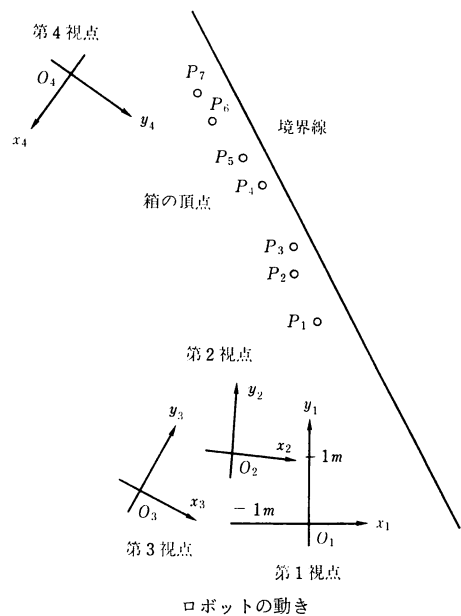
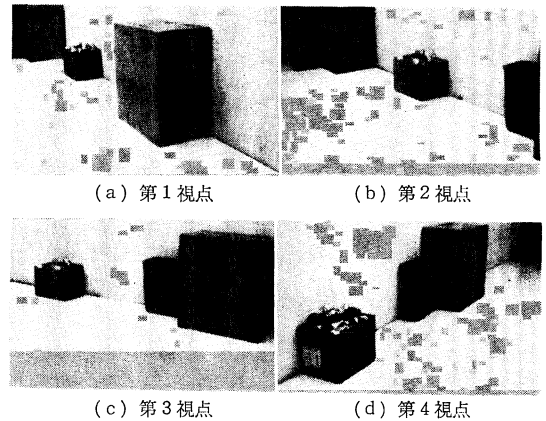
図7 推定された物体の3次元構造⁸⁾図8 移動ロボットの視覚¹⁸⁾

像から、水平に近いときには垂直断面画像から3次元情報を得る。図6に対象物体の画像と、その視点の移動による動画像から生成された、白線の位置での水平・垂直断面画像の例を示す。この断面画像から線分を検出しその傾き、すなわち視差を計算し、物体までの距離を求める。得られた物体の構造を図7に示す。この図では物体の3次元輪郭線を実線で、その3面図上への投影を点線で記している。

5. 知能ロボットへの応用

3次元動画処理の最も大きな応用先の一つは、移動ロボットの視覚システムであろう。ここでは、移動ロボットが安全に移動するために必要な歩行者の検出および追跡¹⁸⁾と、ロボット自身の動きを動画像から推定する方法¹⁹⁾を示す。前節までは動きや構造の一般的な推定法について述べてきた。しかし、移動ロボットの特性や環境の特徴をうまく利用すれば、より簡単な方法で人間の動きやロボットの動きを知ることができる。

図8のような移動ロボットを想定しよう。ロボットは平らな床面(室内や廊下等)上を移動し、その視覚は床面上を一定の角度で見下ろしているとする。このとき、歩行者のつま先の位置は床面上にあることから、画面上でのその投影位置から、歩行者の実際の位置が得られ

図9 移動画像とロボットの動き¹⁹⁾

る。したがって、つま先の画面上での追跡から床面上での歩行者の動きを即座に知ることができる。

一方、ロボットが床面上にあるいくつかの点を観測しつつ移動することにより、ロボット自身の動きを知ることができる。図9のa~dは四つの異なる視点から見た廊下のシーンである。これらの画像すべてに壁と廊下との境界線が見えている。この境界線の画面上での位置からロボットと壁の距離を、また境界線の傾きから壁に対するロボットの向きを知ることができる。さらに、廊下に置いてある特定の箱の位置を観測することにより、ロボットの廊下上での位置を完全に決定することができる。個々の画像から得られたロボットの位置と向きの情報からロボットの移動量が推定される。このようにして

得られたロボットの動きが図9に記されている。

6. あとがき

物体の3次元運動や3次元形状を動画像から推定する手法のいくつかを紹介した。これらの手法を知能ロボットの視覚への応用するためには処理速度の向上が課題である。光コンピュータやニューラルネットワークなどのような新しいアーキテクチャによるコンピュータの出現は処理の高速化のほかに、新たな3次元動画処理手法の開発が期待できよう。

文 献

- 1) C. Cafforio and F. Rocca: "Methods for measuring small displacements of television images," IEEE Trans. Inform. Theory, **IT-22** (1976) 573-579.
- 2) J. O. Limb and J. A. Murphy: "Estimation of velocity of moving images in television signal," Comput. Graph. Image Process., **4** (1975) 311-327.
- 3) B. K. P. Horn and B. G. Schunk: "Determining optical flow," Artif. Intell., **17** (1981) 185-203.
- 4) 尾上守夫, 前田紀彦, 齊藤 優: "残差逐次検定法による画像の重ね合わせ", 情報処理, **17** (1976) 634-640.
- 5) 浅田 稔, 谷内田正彦, 辻 三郎: "運動物体の検出と追跡", 電子通信学会論文誌, **J 62-D** (1979) 395-402.
- 6) S. T. Barnard and W. D. Thompson: "Disparity analysis of images," IEEE Trans. Pattern Anal. Mach. Intell., **PAMI-2** (1981) 333-340.
- 7) H. Miike, Y. Kurihara, H. Hashimoto and K. Koga: "Velocity-field measurement by pixel-based temporal mutual-correlation analysis of dynamic image," Trans. Inst. Electron. Commun. Eng., **E69** (1986) 877-882.
- 8) 山本正信: "連続ステレオ画像からの三次元情報の抽出", 電子通信学会論文誌, **J 69-D** (1986) 1631-1638.
- 9) G. Adiv: "Determining three-dimensional motion and structure from optical flow generated by several moving objects," IEEE Trans. Pattern Anal. Mach. Intell., **PAMI-7** (1985) 384-401.
- 10) 金谷健一: "対応点を用いない物体の運動認識の理論", 情報処理学会論文誌, **27** (1986) 373-382.
- 11) R. Y. Tsai and T. S. Huang: "Estimating three-dimensional motion parameters of a rigid planar patch," IEEE Trans. Acoust. Speech Signal Process., **ASSP-29** (1981) 1147-1152.
- 12) M. Subbarao and A. M. Waxman: "Closed form solutions to image flow equations for planar surfaces in motion," Comput. Vision Graph. Image Process., **36** (1986) 208-228.
- 13) 山本正信: "勾配法による動画像と距離情報からの複数動物体の3次元運動解析", 電子通信学会論文誌, **J 69-D** (1986) 785-793.
- 14) L. Dreschler and H.-H. Nagel: "Volumetric model and 3D trajectory of a moving car derived from monocular TV frame sequences of a street scene," Comput. Graph. Image Process., **20** (1982) 199-228.
- 15) W. E. L. Grimson: "Computational experiments with a feature based stereo algorithm," IEEE Trans. Pattern Anal. Mach. Intell., **PAMI-7** (1985) 17-34.
- 16) Y. Ohta and T. Kanade: "Stereo by intra-and inter-scanline search using dynamic programming," IEEE Trans. Pattern Anal. Mach. Intell., **PAMI-7** (1985) 139-154.
- 17) 大森隆司, 森下 巖: "多視点ステレオ画像を用いた物体検出", SICE 論文集, **18** (1982) 716-722.
- 18) 築山俊史, 白井良明: "連続する画像からの人の動きの検出", SICE 論文集, **19** (1983) 909-915.
- 19) 築山俊史, T. S. Huang: "自律的な車の走行のためのモーションステレオ", SICE 論文集, **22** (1986) 968-975.