

## 無人走行車のためのモーションステレオ

## Motion Stereo for Navigation of Autonomous Vehicles

築山俊史 (電子技術総合研究所) Thomas S. Huang (Univ. of Illinois)  
(Toshifumi Tsukiyama, Electrotechnical Lab.)

ABSTRACT- In order to navigate, vehicles need to deduce their movements in given environments. We describe a methods for wide-angle motion stereo to aid such navigations. Our approach uses vertices of objects observed in a scene as features to be matched in the two images. The matching is done in the 3-d world with the guidance of a base line which can be seen in both images. The candidate vertices for matching are obtained from analyzing (1)The edge maps of input images using a junction type table and (2)The range information of the segments which form the vertices. Range information is obtained by inverse perspective transformation. Some experimental results are shown.

## 1 はじめに

本論文では、撮像位置がかなり離れた2枚のTV画像データから、そのカメラの動きを推定するモーションステレオを提案する。提案するモーションステレオは、人工の環境下で自律走行する車のビジョンシステムへの応用を想定したものである。

自律的な車の走行には、走行環境からの情報をもとに、自分の移動量を推定する機能が要求される。TV画像から移動量を推定する方法として、既知図形を使う方式[1]や2組のステレオ画像を使う方式[2]などが提案されている。汎用性のためにはシーンで一般に見られるような特徴点を用いるのが望ましい。しかし特徴点の抽出や位置の計測のための処理にかなりの計算時間を必要とする。そのため走行中の撮像回数ができるだけ少ないのが望ましい。しかし画像のサンプリング間隔が長くなるにつれて、対象物体の見かけの大きさや位置が大きく変わり、画像間で特徴点の対応付けが困難になる問題が生じる。従来の画像のマッチング手法の多くは画像上での対象物体の見かけの大きさや位置の変化が小さい場合を扱っている[3-5]。見かけの変化が大きくなる場合、対象物体に固有な構造情報を手掛かりに使うのが有効である。しかし2次元の構造情報たとえば画像から得られた対象物体の線画だけでは、画像間の対応付けにあいまいさがある[6]。本論文では、画像間の見かけの変化が大きい場合でも有効なように、シーンの3次元的な構造情報を

使って画像間の対応付けを行い、カメラの動き(即ち車の移動量)を求める方法を提案する。シーンの構造の解析には画像の逆透視変換から得られる距離情報を使う。提案する方式の概略は次節で述べるが、ここでは車の走行空間として建物内の廊下を想定し、またTVカメラは、移動前後で同一側の壁面を向いているものとする。第3~5節で画像からシーンの構造情報を抽出する方法の詳細を述べる。第6節では、マッチングの方法を述べ、第7節では実画像を使った実験結果について考察する。

## 2 処理の概要

図1にTV画像の入力方式とカメラの座標系を示す。TVカメラは $h$ の高さから見込角 $\theta$ でシーンの画像を取る。座標系 $(x, y, z)$ は移動するカメラに固定され、 $x$ と $y$ 軸は床面にあり、カメラの光軸は $yz$ 平面にあるものとする。図2に提案するモーションステレオの原理を示す。カメラは時刻 $t_1$ に $Q_1$ の位置から、シーン内の床面上の直線 $g$ と点 $P$ を観測し、その位置を求める。カメラは車と共に移動し、時刻 $t_2$ に $Q_2$ の位置から、同じ直線と点を同様に観測する。混乱をさけるために、図2では $z$ 軸は省略されている。このときカメラの回転および並進成分と観測された点 $P$ の位置の間に次のような関係がある。

$$\begin{bmatrix} x_1 & y_1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -t_x & -t_y & 1 \end{bmatrix} \begin{bmatrix} \cos(-\phi) & -\sin(-\phi) & 0 \\ \sin(-\phi) & \cos(-\phi) & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} x_2 & y_2 & 1 \end{bmatrix}$$

ここで、 $(x_1, y_1)$  と  $(x_2, y_2)$  はそれぞれの座標系での点Pの位置を表わす。 $(t_x, t_y)$  はカメラの並進成分で  $\phi$  はz軸回りの回転角を表わす。図2より、カメラの回転角 ( $\phi$ ) はそれぞれの座標系からみた直線gの方向 ( $\alpha_1, \alpha_2$ ) により決まる。カメラの並進成分

$(t_x, t_y)$  はこの回転角と点Pの座標値から計算できる。ここでは直線gのことを基準線、点Pを特徴点と呼ぶ。提案する方式は同一の基準線と特徴点に相当する物を2枚の画像の中で見つけ、その位置の情報からカメラの動き、即ち車の移動量を求めるものである。

ところで建物の廊下のような人造物は、通常、水平な平面(床面)と垂直な平面群(壁、柱、戸など)できている。壁と床面との間には長い直線の境界線ができ、また他の垂直な平面群と床面との間には凸の頂点が多く見られる。これらはシーンのどの場所からでも観測できるので、この境界線を基準線として、また凸の頂点を特徴点として使う。この境界線や凸の頂点は床面にあるので位置は画像の逆透視変換から求める[7]。即ち、図1で示された平面上の点Pの位置はその画像アドレス(Q)、カメラの高さ(h)、見込角( $\theta$ )などから計算できることを利用する。

全体の処理の流れは次のようになる。①まず画像から画素間の明るさの変化量をもとに、境界線と凸の頂点の稜線の候補になるエッジ部分を抽出し、エッジをセグメント化する。②床面上にあるセグメントで、その距離が廊下の最大幅をこえない物を境界線の候補とする。このなかで長さが一定値以上で、その位置が最も遠方にあるセグメントを境界線とする。③セグメントの位置と接続関係から凸の頂点を見つける。ここでは頂点を三面頂点と仮定する。④頂点を形成するセグメントの長さ、角度などを使い、画像間で同じ形状の頂点の対を作る。境界線は画像間で共通であり、この境界線にそった頂点の順序と頂点間の相対的な距離は、見る方向が変わっても不変である。頂点の対のなかでこの制約条件を満足する頂点の対の組み合わせを作る。この組み合わせが画像間で対応付けされた頂点になる。⑤対応付けされた頂点と境界線の位置から前述の方式でカメラの動きを計算する。

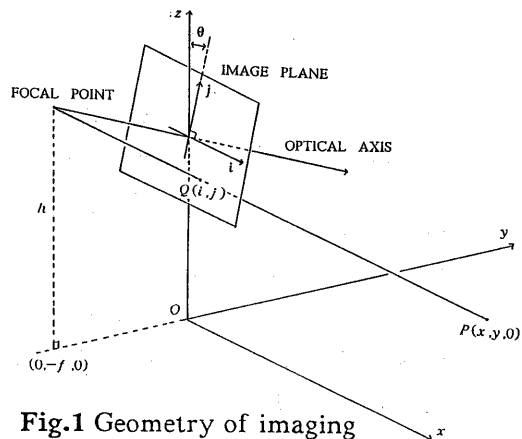
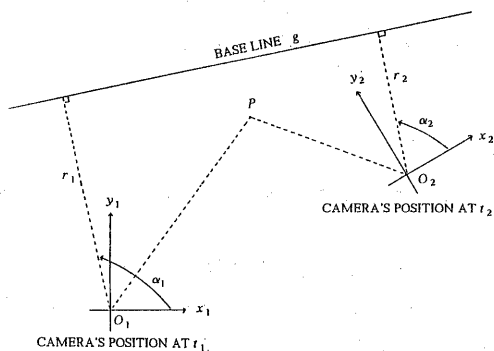


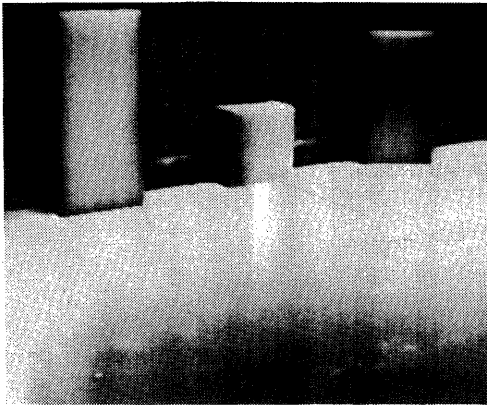
Fig.1 Geometry of imaging a scene by a TV camera.



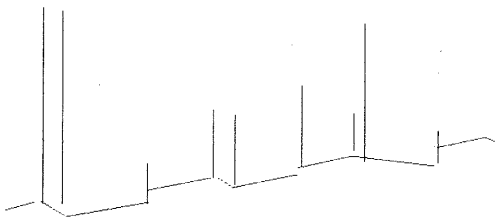
### 3 セグメントの抽出

この節では、入力画像からエッジ部分を抽出し、これからS、HとVセグメントを決定する方法について述べる。SとHセグメントは床面と垂直な平面との間の稜線に対応する線分で、Sセグメントは画像上で傾いている線分のことである。Hセグメントは水平な線分を意味する。Vセグメントは垂直な平面間に見える稜線に対応している。廊下は天井から十分な照明で一律に照らされているものとする。このとき、もし床面と同じ反射率をもつ平面がその廊下に垂直に立っていたとすると、床面からある程度の高さまで、その垂直面の輝度は床面のそれより小さく観測される。ここでさらに、床面は比較的反射率の高い平面できているとする。照明の仮定から、床面はTV画像の中央下部に位置する明るい領域になり、S、Hセグメントの候補となる画素は、明るさが大きく変化し、エッジの方向が画像上で下方へ向う。またVセグメントの候補点は、TV画像中で垂直なエッジ部分である。

エッジの検出にはPrewitt のオペレータで空間微分をおこない、各画素のエッジの強さと方向を求める。セグメントごとに条件にあう画素を選ぶ。次に、セグメントの種類ごとにエッジの候補点となった画素の連結性（4連結）を調べ、近接する画素どうしグループ化する。セグメントは廊下を構成している平面間のできる稜線に対応しているの、画素数が少ないグループ、あるいはエッジの方向が分散しているグループはセグメントの候補から削除される。残りの各グループは一定の長さの線状領域になるので、この領域の長軸方向の中心線をそのグループを代表するセグメントとする。このセグメントにラベル付けをし、それぞれの種類ごとのリストに登録する。図3(a) は入力画像の例である。図3(b) はこの節で述べた処理によって得られたセグメントである。



(a) Original image.



(b) Overlay of the extracted segments.

Fig.3 Example of extraction of edge lines.

#### 4 廊下の境界線の推定

この節では、距離情報を使い、SとHセグメントのグループから廊下の境界線を推定する方法について述べる。廊下の境界線というのは、廊下の柱などの部分的なでっぱりを取り除いたとき、側壁に相当する仮想的な垂直の大きな平面と、床面とが交わってできる稜線のことである。この稜線はTVカメラの方向を決めるための基準線としてだけでなく、次節で述べるセグメントの接続関係から頂点を見つけるときのために、そして2つの画像間での頂点の対応付けのガイドとして使われる。

前節では明るさの変化をもとにセグメントを抽出したが、得られたS、Hセグメントのグループには床面との稜線だけでなく、物体の上辺にあたる稜線も含まれている可能性がある。そのためセグメントの廊下での位置を計算して、床面上の稜線である可能性が小さいセグメントをリストから削除する。まず、セグメントは全て床面にあるとみなし、それぞれの線分の両端を延長した直線を考える。そして、TVカメラの位置を原点とする座標系を床面上に考え（図1参照）、原点からそれぞれの直線までの距離を計算する。距離は逆透視変換を使い求める。距離がかなり大きい場合、セグメントは床面ではなく、その上方にある（物体の上辺にあたる稜線）と考えられるから、この距離が一定値（廊下の最大幅）を越るものをセグメントのリストから除く。そして残りのS、Hセグメントのグループから廊下の境界を推定する。

境界線はTVカメラの方向により画像中では正か負の傾きを持つ直線、あるいは水平な直線になる。ここで“正”の傾きというのは画像座標系で直線の傾きが正になるものをさす。廊下の境界線の近辺には柱などの物体と床面によりいろいろな方向に稜線ができるが、境界線に沿う方向に多くのセグメントがみられる、そのためSセグメントのグループを傾きの正負により2つに分割し、各グループごとにセグメントの長さを合計する。そしてその値が最大になるグループを境界線の候補とする。このグループ内の直線のうち、互いに平行で、その両者の距離が短いのであれば、もとの2つのセグメントは同じ直線上にあるとみなし、そして、両者を併合して1つの直線に置き換える。直線上にあるセグメントの線分の長さの合計が一定値以上で、その距離が最大の直線を廊下の境界線とする。

### 5 セグメントの解釈

廊下を3面頂点世界と仮定すると、画像に現れるセグメントの接続点のタイプは図4のようなになる[7]。接続点Pの物理的形狀は、W型とL型では床面にある凸の頂点、Y型では床面と2つの垂直な平面による凹の頂点になる。T型では稜線の一部が垂直な物体で隠されている。この節ではセグメントの接続関係から接続点のタイプを決める方法について述べる。

セグメントの探索は最も確実なセグメント、即ち廊下の境界線になるセグメントから、隣接しているセグメントを順次見つけていく。セグメントの探索は、図5のような限定された領域で行う。この領域内にセグメントの端が見つかれば、両者は接続しているものとして、それぞれのセグメントを延長して交点を求める。この交点を接続点とする。Vセグメントは実際よりかなり短くなることが多いため、探索範囲を上方向に長くとってある。

すべてのセグメントの接続点を見つけたあと、形状は次の方法で決定する。Vセグメントの下端に2つのSセグメント（あるいはHとSセグメント）が接続している場合、または2つのSセグメント（あるいはHとSセグメント）だけの接続点は、SやHセグメントの床面上の位置関係から、その形状をW型かY型とする。VセグメントとSセグメント（あるいはHセグメント）が接続している場合、完全なセグメントである保証がないので、これだけから形状の判定はできない。たとえば図6に示す例では2通りの可能性がある。ここではVセグメントに別のS（あるいはH）セグメントが接続しているかを調べ、図7のように左右に1つずつセグメントがあれば、 $P_1$ をL型、 $P_2$ をT型とする。これ以外のVセグメントとの接続点や、他との接続が見つからなかったS（あるいはH）セグメントの端点は孤立点とする。図8は、図3(b)のセグメントの接続関係を表わす。接続点のマークは次のような意味を持っている。+ : W型、□ : Y型、■ : T型、◇ : L型、▲ : 孤立点、H : 画像の境界点。

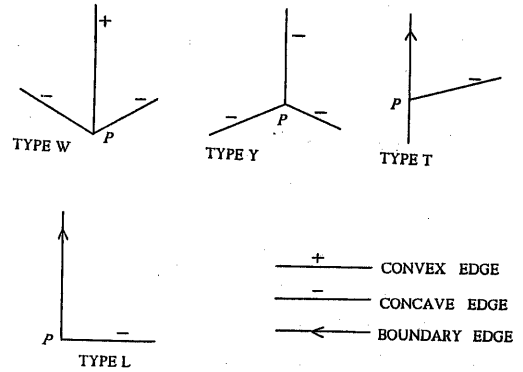


Fig.4 Junction types.

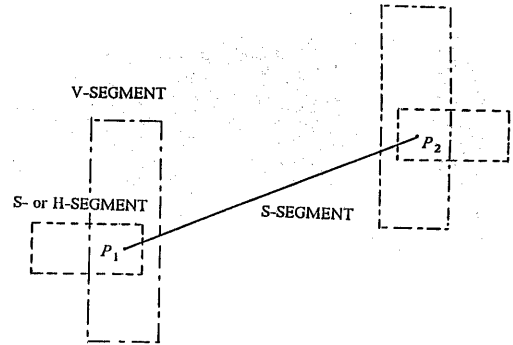


Fig.5 Searching areas for junction points.

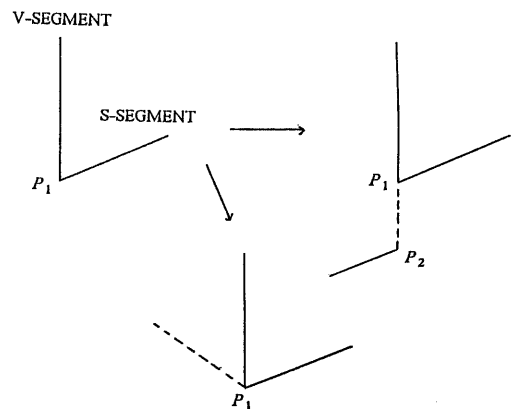


Fig.6 Ambiguity of junction types.

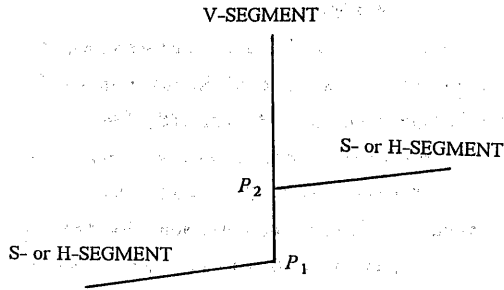


Fig.7 V-segment with two junction points.

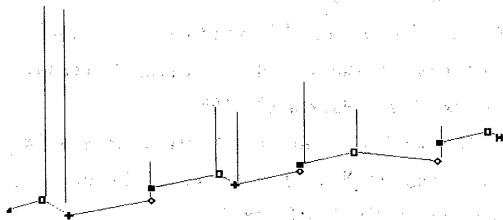


Fig.8 Vertices in Fig.3(b).

## 6 頂点の対応付け

この節では画像間の頂点の対応付けの方法について述べる。対応付けは各画像から得られた頂点の3次元世界での特徴量を使って行なう。

前節で得られた頂点のうちW型とL型の頂点は凸のため、撮像位置が変わっても見え隠れする可能性が少ない。これを対応付けのため特徴点として使う。すべてのセグメントが完全な形で抽出されている保証がないので、孤立点とラベル付けされた点も凸頂点である可能性がある。そのため境界線から一定値以上はなれている孤立点も特徴点として使う。画像間での頂点の対応付けのために次のような特徴量を使う。

- ①境界線から頂点までの距離
- ②頂点を形成する左側のS（あるいはH）セグメントの長さ
- ③頂点を形成する右側のS（あるいはH）セグメントの長さ
- ④2つのセグメントによりできる角度

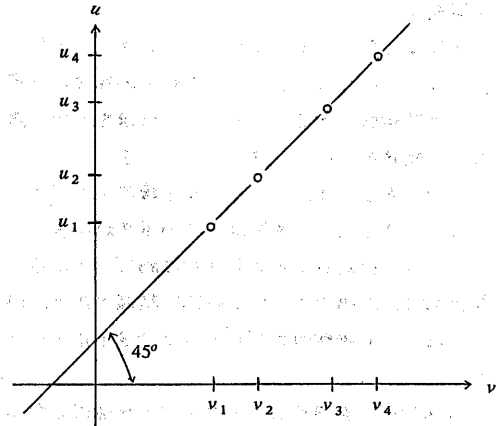


Fig.9 Mapping matching pairs.

ここで左側のセグメントというのは観測者からみて左側にあるものをさす。またセグメントの長さ及び角度は3次元世界での量を割り当てる。ところでL型や孤立点の場合、1つのセグメントが隠されるか消失している。この場合頂点から境界線までの距離をセグメントの長さとする。また頂点がW型の場合でも、もし境界線に垂直なセグメントが小さな物体に隠され、セグメントの1部が欠けていたら、頂点から境界線までの距離をそのセグメントの長さとする。これは画像間で見える方向が大きく変化してもW型とL型の間での対応付けが正しく行なわれるようにするためである。

対応付けはそれぞれの画像から頂点を1つずつ選び、それぞれの特徴量の差が一定値以下であれば、2つの頂点是对应付けの候補対にする。候補対から、最終的な対応付けは次のように行なう。ここでは、境界線にそった頂点の順序ならびに頂点間の相対距離はTVカメラの位置が変わっても不変であることを利用する。まず対応点の候補は図9に示すような座標系(u,v)にプロットする。u軸は1番目の画像中の境界線に対応し、v軸は2番目の画像中の境界線に相当する。それぞれの座標軸の原点は、境界線の画像での左端の点である。uとvの座標値は頂点をそれぞれの境界線へ垂直に投影したときの原点からの距離である。上記の制約条件から正しい対応付けは、傾き1の直線上に在る。それ故に、この傾きを持つ直線群の中で、対応点の数が最も多いものを取り出し、その直線上にある候補対を正しい対応点とする。

## 7 実験結果

ここで提案した方式を様々なシーンを使って検証した。図10(a) - (d) は同じシーンを4つの異なる位置から取ったTV画像である。このシーンは研究室の一隅に箱などの物体を置いて、廊下のシーンをシュミレートしたものである。床面はアイボリ系統のプラスチックタイルでできており、照明は天井の蛍光灯による。TVカメラの高さは約1mでその見込角は約17°である。TV画像は512X512画素で各画素は256階調である。画像データはVICOM画像処理プロセッサで量子化され処理された。

図11は図10(a)を処理し、セグメントを抽出した結果である。マークの付いた頂点は対応付けの候補点を表わしている。図12は図10(a)と(b)との間での対応がとれた頂点を示している。さらに図13は図10(a)と(c)との間での対応付け、そして図14は図10(a)と(d)との間での対応付けの結果を示している。

提案した方式ではカメラの移動量は1対の頂点から決定できる。しかし画像のノイズなどの影響で、頂点の位置は誤差を含んでいる可能性がある。そのため得られたすべての対応点ごとにカメラの移動量を計算し、その平均値を最終的な移動量とした。図15は図10(a) - (d)間でのカメラの移動量を示している。座標系の原点を図10(a)でのTVカメラの位置にとってある。各y軸の方向はカメラの視軸方向を示している。これらの結果から見る方向が大きく変化しても対応付けが有効にできていることがわかる。

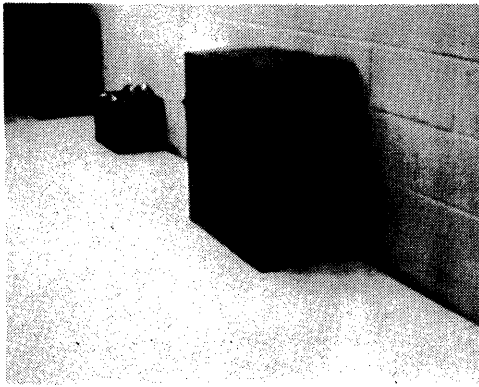
## 8 おわりに

人工の環境下において車が自律的に走行するためのモーションステレオについて提案した。画像間の対応付けにはシーン内の物体の頂点を特徴点として使った。対応付けにはシーンの中で見られる基準線をガイドとして使った。物体の頂点は入力画像のエッジマップを接合テーブルを使い解析し、求めた。距離の情報は逆透視変換により求めた。いくつかの実験結果について述べた。

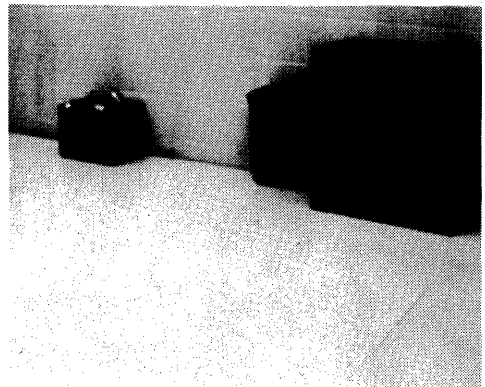
対応付けにはシーンの3次元の構造情報を使っている。そのためこの対応付けの方式は車が移動しながら走行空間のモデルを作成するのにも応用が可能である。

## REFERENCES

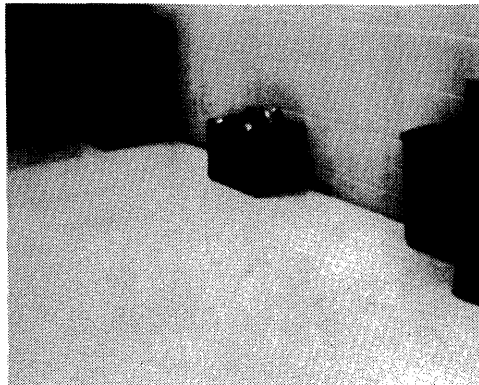
- [1] I. Fukui, "TV Image processing to Determine the Position of Robot Vehicle", Pattern Recognition Vol. 14 p101-109, 1981.
- [2] H. P. Moravec, "Rober Visual Obstacle Avoidance", Proceedings of 7th IJCAI, 1981.
- [3] R. Nevatia, "Depth Measurement by Motion Stereo", Computer Graphics and Image Processing Vol. 5 p203-214, 1976.
- [4] S. T. Barnard & W. B. Thompson, "Disparity Analysis of Images", IEEE Transaction of PAMI Vol. 2 p333-340, 1980.
- [5] M. J. Hannah, "Bootstrap Stereo", Proceedings of Image Understanding Workshop (at College Park, Maryland), 1980.
- [6] C. Thorpe & S. Shafer, "Correspondence in Line Drawing of Multiple Views of Objects", Proceedings of 8th IJCAI, 1983.
- [7] R. O. Duda & P. E. Hart, "Pattern Classification and Scene Analysis", John Wiley & Sons, 1973.



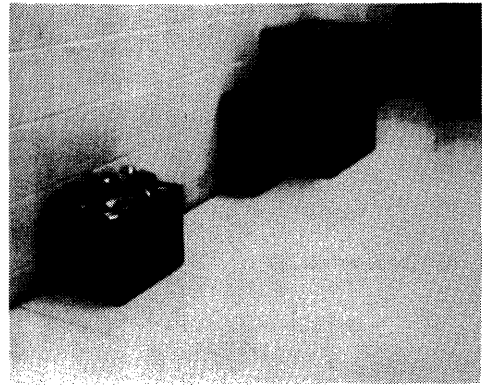
(a) First image.



(b) Second image.

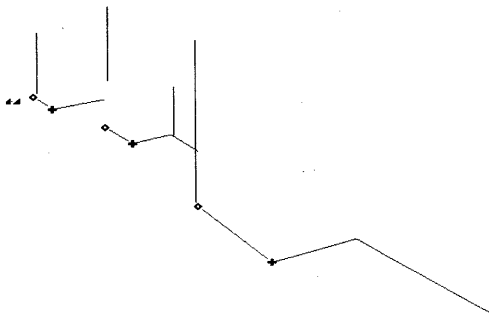


(c) Third image.

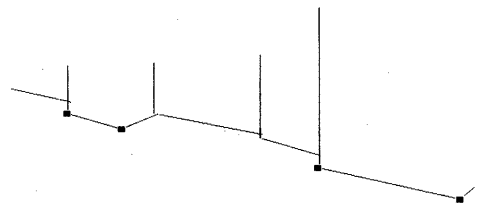


(d) Fourth image.

**Fig.10** Four different images of the same scene.



**Fig.11** Vertices in Fig.10(a) for matching.



**Fig.12** Matched vertices in Fig.10(b).

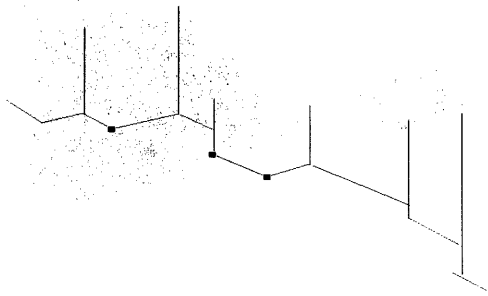


Fig.13 Matched vertices in Fig.10(c).

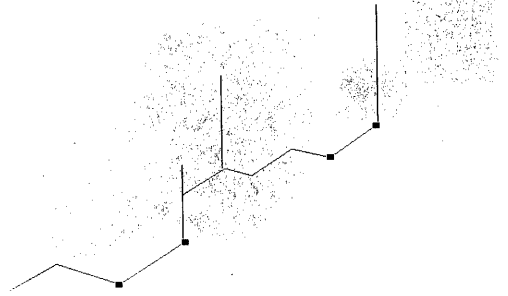


Fig.14 Matched vertices in Fig.10(d).

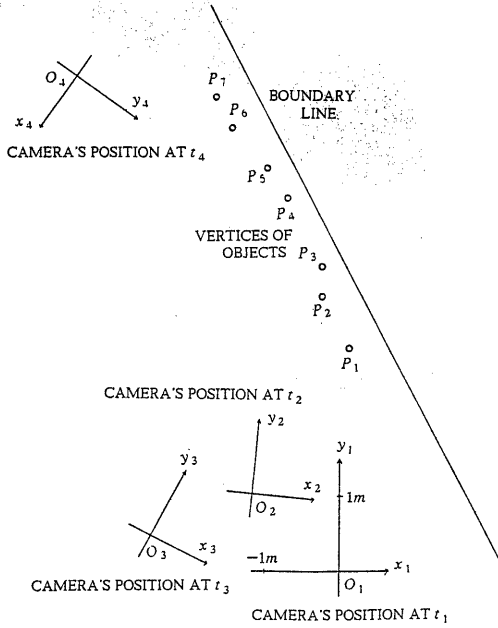


Fig.15 Camera movements  
in the scene of Fig.10(a)-(d).