

アクティブビジョンにおける箱型物体の認識の検討

山内俊明¹・佐藤純²・増田大樹³・関靖夫¹

¹ 情報メディア学科 (yama@ic.kanagawa-it.ac.jp)

² 情報工学科 (sato@sek.ic.kanagawa-it.ac.jp)

³ 三菱ウェルファーマ株式会社

Investigation of Recognition for Box-shaped Object

Tosiaki YAMANOUCHI¹, Jun SATO², Daiki MASUDA³, Yasuo SEKI¹

Abstract

The purpose of this study is to establish an active vision algorithm to control a robot arm for object classification operation. In our previous work, a method of selecting an object to hold by the robot arm was proposed, and was experimentally examined by using computer generated images. As the second step of this study, this paper describes a method of estimating value of the three dimensional coordinate of vertex of object by using real images. The experimental results show that the proposed method can estimate a box-shaped object geometry with a high degree of accuracy.

Key Words: Robot Vision, Active Vision, Object Classification, Stereo Vision

1. はじめに

カメラを人間の目のように状況に応じて動かして情報を取得するアクティブビジョンという研究分野がある。この分野は、最近脚光を浴びている家庭もしくはオフィスで作業を行う汎用ロボットの視覚システムの基礎となるものであるが、撮像条件などに左右されやすいためにヒューリスティックになりがちである。本研究では、撮像条件に制約を与えない一般的な環境におけるアクティブビジョンを利用した荷物仕分けロボットの処理・制御アルゴリズムの開発を目的としている。カメラを搭載した自律型ロボットに荷物の仕分けを行わせるためには、与えられた視覚情報から荷物の形状・大きさ・位置を認識するとともに、より精度の高い情報や不足している情報を得るためのカメラ移動先の決定が必要となる。

我々は先行研究^[1]において、複数の物体からロボットアームが最初に把握する物体をカメラ情報より決定する処理について検討を行い、その一部を計算機で生成したシミュレーション画像（簡単のために荷物はすべて箱型とした）により検証した。具体的には、

- 1) 初期画像から3面頂点（3面に含まれる頂点）を同定し、画像中で最上部にある3面頂点を対象物体とする。
- 2) カメラ微動によって得られた2枚の画像より、当該頂点及び隣接頂点の3次元座標を算出する。

- 3) カメラから向かって当該頂点の右側に位置する隣接頂点を回転中心とする回転運動をカメラに与え、対象物体の反対側面の視覚情報を取得する。

- 4) 当該視覚情報より、ロボットアームで把握するための空間的余裕が存在するか否かを得る。

というものであった。今回は第二段階として、カメラを微動して得られる2枚の実画像から物体の頂点の3次元座標の推定を行い、その精度を検討することにした。簡単のために、撮影対象とする物体は箱型とするとともに、画像中には単一の物体しか含まれないものとした。

具体的な処理の流れは次の通りである。まず、処理対象のフルカラー画像をグレースケール画像に変換し、エッジ検出と細線化を行う。その後、直線ハフ変換により直線を同定するとともに、その交点を算出する。そして、この得られた全交点から実質交点（実際にエッジの頂点となっているもの）を抽出し、2枚の画像で頂点の対応付けを行う。最後に、ステレオ法により対応する頂点の2次元座標から3次元座標を算出する。

今回と先行研究との違いは、実画像を用いる点にある。計算機シミュレーションで生成した画像は、エッジの強度も均一であるために両画像内の頂点の2次元座標を正確に求めることができる。しかし、実際に光学的に撮影された画像は、照明や物体の質感などの影響を受けて、画像処理・認識の過程で雑音が入ってしまう。そのた

め、直線や交点の同定に用いられる閾値にはいくらかの冗長性が要求される。また、撮影に用いるカメラの焦点距離も、予備実験を行うことにより実際に算出を行ったのでこれについても述べる。

2. ステレオ法による物体頂点の三次元座標推定

2.1 処理の概要

三角測量法の原理により、同じ頂点を別の視点から撮影した複数の画像内の点として対応させることができれば、当該座標の三次元座標を求めることができる。しかし、複数の画像に写った任意形状物体の特徴点の三次元座標を推定することは容易なことではない。その原因としては、物体の形状が未知である場合には複数の画像間での特徴点の対応付けが非常に困難であるからである。また、対応付けができたとしても、実画像の場合には検出した特徴点の二次元座標に推定誤差が発生するため、ステレオ法によって算出される三次元座標の精度は期待できない。今回は扱う対象を箱型物体、すなわち直方体に限定し、さらにカメラの移動を水平方向にわずかな距離とすることによって、頂点の対応付けの問題を回避することにした。

2枚の画像から、物体の頂点の三次元座標を算出するまでに必要な処理を図1の流れ図に示す。

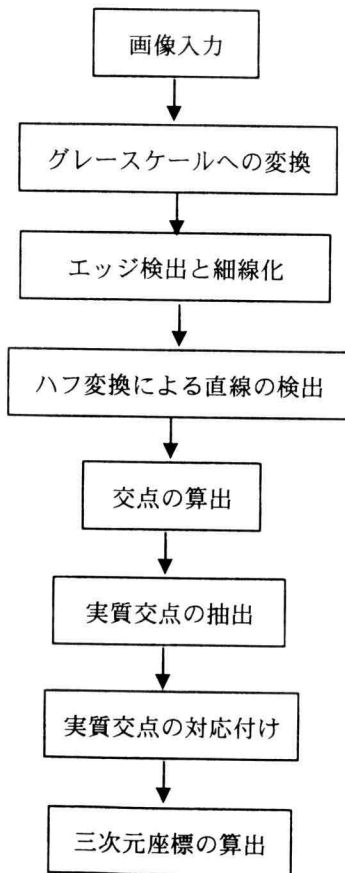


図1 物体の頂点三次元座標の推定手法の流れ図

画像中の箱型物体の頂点が物体の稜線の交点となっていることから、まずエッジを検出してハフ変換を行うことにより稜線の推定を行う。次に、推定した稜線相互の交点を求め、その中から実際にエッジの交点となっている実質交点（これが実際の頂点对応する）を抽出する。そして、二枚の画像間で対応付けた実質交点から、ステレオ法により当該頂点の三次元座標を算出するのである。

2.2 グレースケールへの変換

エッジ検出の前処理という位置づけで、入力されたカラー画像をグレースケール画像に変換する。変換方法は、当該ピクセルの色情報をRGB形式からHSV形式に変換し、HSV形式のS (Saturation; 彩度) を0~255の256段階に変換してグレースケールとして用いるというものである。

2.3 エッジ検出と細線化

ガウス平滑型微分処理をグレースケール画像に適用し、エッジ強度が閾値を越えたピクセルをエッジとする。なお、この場合の閾値はエッジの最大強度に対する割合として与えることにした。その後、細線化処理を行った。

2.4 ハフ変換による直線の検出

検出した全エッジについて、対応する直線ハフ空間上の点に1票を投じる。投票終了後、得票数が閾値を上回り、かつ極大値となっていて、既に検出済みの直線からある距離（閾値として設定する）以上離れている点を直線として検出する。

2.5 交点の算出

検出した全ての直線相互の交点を算出する。ただし、画像の外で直線が交わる場合には、その交点は排除することにした。ハフ変換によって検出した直線の傾きやy切片には、推定誤差が混入している。そのため、本来は3つの直線が1点で交わる場合に、図2に示すように2直線の交点が3点あるような結果になってしまうことがある。この問題を解決するために、次の2つの条件が満たされた場合には2交点の重心を新しい重心の位置として、2交点を1交点に融合するという後処理を施すことにした。

- 1) 交点の位置が一定の距離（閾値を設ける）以内にある。
- 2) 両交点を通る直線に同一の直線がある。

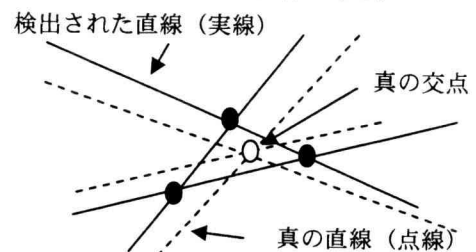
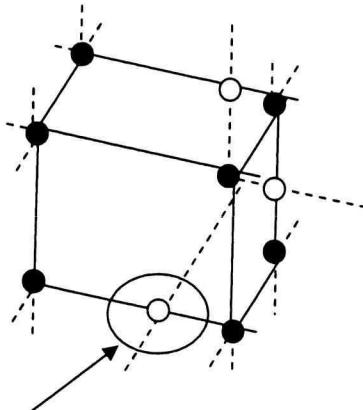


図2 誤差により交点算出に影響がある場合

2.6 実質交点の抽出

求めた交点の内、交点の周囲ある一定の範囲内に当該交点を通ることになっている全ての直線に対応したエッジが存在するものだけを、実質交点として抽出する。抽出の様子を図3に示す。



円内に点線方向のエッジは無いため、実質交点とはならない

図3 実質交点の抽出の様子

2.7 実質交点の対応付け

今回は、カメラの移動を水平方向に限定しているため、他方の画像中の対応する実質交点を探索する場合は、同じ高さを調べればよい。しかし、実質交点の推定座標には誤差が含まれており、必ずしも正確に同じ高さに対応する実質交点が存在するとは限らない。そのため、閾値を設けて一方の画像中の実質交点と同じ高さ周辺にある実質交点を探索の対象とする。

2.8 三次元座標の算出

左右画像の実質交点の座標をそれぞれ (i_l, j_l) , (i_r, j_r) とする。また、対応するカメラの撮像面上の座標をそれぞれ (x_l, y_l) , (x_r, y_r) とする。さらに、撮像面の幅を w_s , 高さを h_s , 画像の幅 (ピクセル数) を m , 高さを n とする。この時、以下の4式が成立する。

$$x_l = \left(-\frac{1}{2} + \left(i_l + \frac{1}{2} \right) \frac{1}{m} \right) w_s \quad (2.1)$$

$$x_r = \left(-\frac{1}{2} + \left(i_r + \frac{1}{2} \right) \frac{1}{m} \right) w_s \quad (2.2)$$

$$y_l = \left(\frac{1}{2} - \left(j_l + \frac{1}{2} \right) \frac{1}{n} \right) h_s \quad (2.3)$$

$$y_r = \left(\frac{1}{2} - \left(j_r + \frac{1}{2} \right) \frac{1}{n} \right) h_s \quad (2.4)$$

左右の画像を撮影した時のカメラ位置の光軸は水平に距離 b だけ離れており、カメラの焦点距離は f とする。これらの位置のレンズ中心を結ぶ軸 (基線) を x 軸とし、レンズ中心の中点を原点とした時の、物体の三次元座標を (x, y, z) とする。この時、以下の3式が成立する。

$$\frac{x_l}{f} = \frac{x + b/2}{z} \quad (2.5)$$

$$\frac{x_r}{f} = \frac{x - b/2}{z} \quad (2.6)$$

$$\frac{y_l}{f} = \frac{y_r}{f} = \frac{y}{z} \quad (2.7)$$

式 (2.5), (2.6), (2.7) を連立させ、 x, y, z について解くと以下ようになる。

$$x = \frac{b}{2} \cdot \frac{x_l + x_r}{x_l - x_r} \quad (2.8)$$

$$y = \frac{b}{2} \cdot \frac{y_l + y_r}{x_l - x_r} \quad (2.9)$$

$$z = \frac{bf}{x_l - x_r} \quad (2.10)$$

式 (2.8) ~ (2.10) と式 (2.1) ~ (2.4) を連立させることにより、左右画像の実質交点の座標と物体の三次元座標の関係が以下のように導出される。

$$x = \frac{b}{2} \cdot \frac{(-m + i_l + i_r + 1)}{i_l - i_r} \quad (2.11)$$

$$y = \frac{b}{2} \cdot \frac{h_s}{w_s} \cdot \frac{m}{n} \cdot \frac{(n - (j_l + j_r + 1))}{i_l - i_r} \quad (2.12)$$

$$z = -\frac{1}{i_l - i_r} \cdot \frac{m \cdot b \cdot f}{w_s} \quad (2.13)$$

3. 焦点距離の算出

組み合わせレンズを用いたカメラの焦点距離を求めるために、以下のように物点・レンズ中心・撮像部の位置関係を定義する。

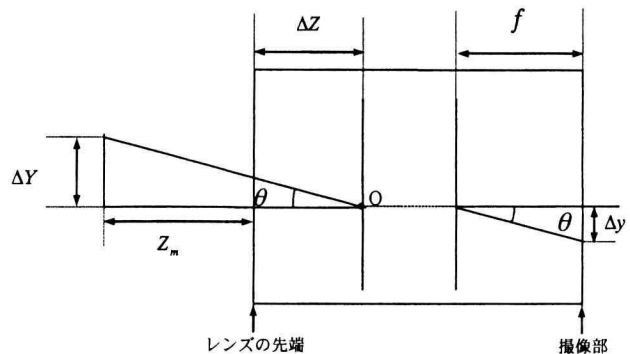


図4 物点・レンズ中心・撮像部の位置関係

焦点距離を算出するためには、レンズ先端から物点までの距離 Z_m を変化させ、その時の画像上の対応するピクセルの座標を記録しておく必要がある。

図4より、比の関係式を以下のように得ることができる。

$$\Delta Y : Z_m + \Delta Z = \Delta y : f \quad (3.1)$$

これを変形して、次式を得る。

$$\frac{\Delta Y}{Z_m + \Delta Z} = \frac{\Delta y}{f} \quad (3.2)$$

撮像部と画像の高さと横幅 (ピクセル数) の関係を図5

に示す。

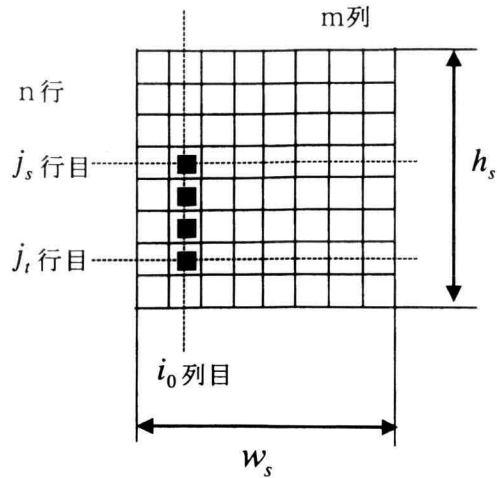


図5 撮像部と画像の幅と高さの対応

$\Delta j = j_s - j_t$ と定義し、下式が成立すると、

$$|\Delta y| = |\Delta j| \cdot \frac{h_s}{n} \quad (3.3)$$

以下の式が成り立つ。

$$|\Delta Y| = (Z_m + \Delta Z) \cdot \frac{h_s}{f \cdot n} \cdot |\Delta j| \quad (3.4)$$

$\alpha = \frac{h_s}{f \cdot n}$, $\beta = \alpha \cdot \Delta Z$ と定義すると、式 (3.4) は次のようになる。

$$\frac{|\Delta Y|}{|\Delta j|} = \alpha \cdot Z_m + \beta \quad (3.5)$$

さらに $\varepsilon = \frac{|\Delta Y|}{|\Delta j|}$ と定義し、レンズ先端から物点までの距離

を変えて測定したデータによる ε の組み合わせを

$$\varepsilon_1 = \alpha \cdot Z_{m1} + \beta \quad (3.6)$$

$$\varepsilon_2 = \alpha \cdot Z_{m2} + \beta \quad (3.7)$$

とし、これを連立させると次のようになる。

$$\alpha = \frac{\varepsilon_1 - \varepsilon_2}{Z_{m1} - Z_{m2}} \quad (3.8)$$

$$\beta = \varepsilon_1 - \alpha \cdot Z_{m1} \quad (3.9)$$

よって、 $\alpha = \frac{h_s}{f \cdot n}$ と式 (3.8) を連立させることにより、

焦点距離 f は次式で求めることができる。

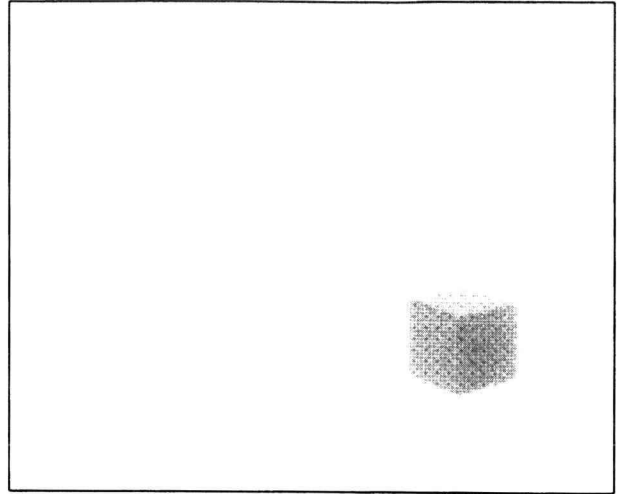
$$f = \frac{h_s}{\alpha \cdot n} \quad (3.10)$$

4. 実験

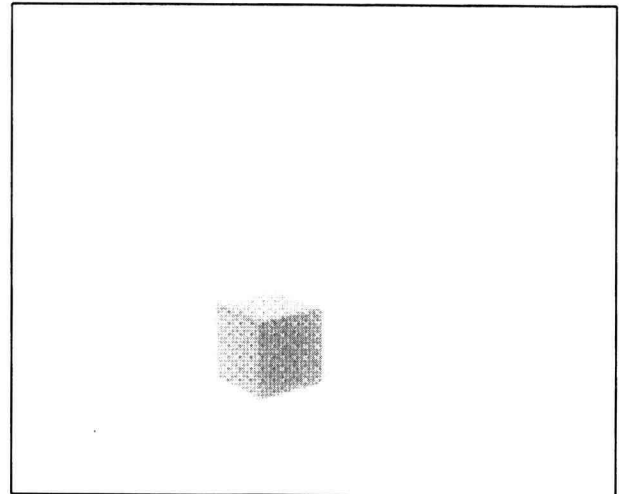
まず、使用するカメラの焦点距離を求めるために、レンズ先端と物点との距離を変化させて、撮影を行った。得られたデータを前章で導出した式に代入したところ、焦点距離として 27 [mm] という値を得た。

次に、ステレオ法によって物体頂点の三次元座標を推定するために必要な、画像の撮影を行った。具体的には、

机の上に一辺が 45 [mm] である立方体（プラスチック製の子供用の積み木で、安全のために多少角が面取りされている）を置き、物体に向かって水平右方向に 100 [mm] 移動する前後に撮影を行った。この時、得られた画像を図 6 に示す。



(a) 移動前に撮影した画像 (左画像)

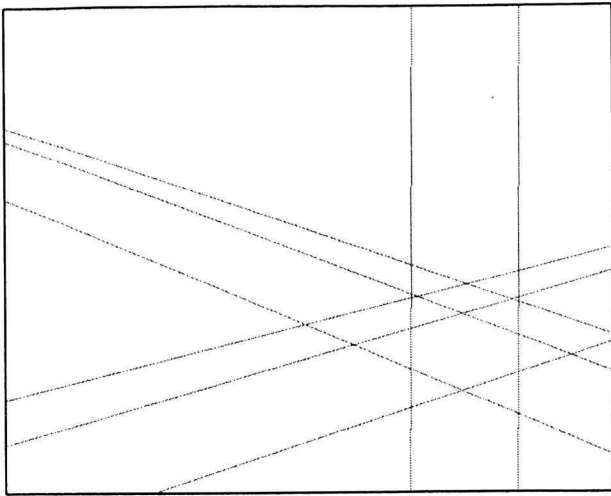


(b) 移動後に撮影した画像 (右画像)

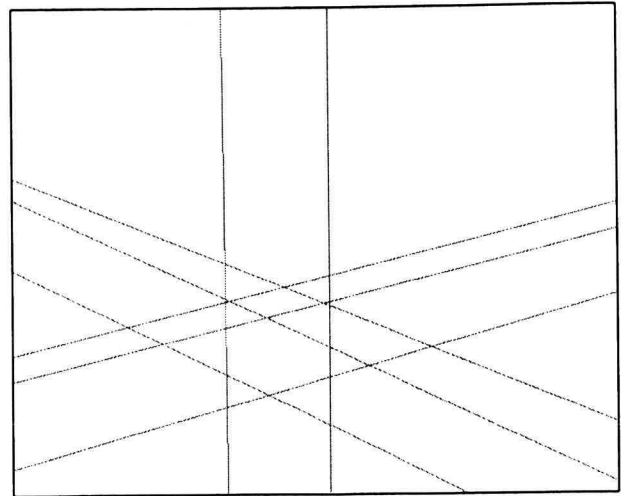
図6 処理対象画像

水平にカメラを移動したので、物体は画像中のほぼ同じ高さに写っていることが伺える。

次に、これらの画像をグレースケールに変換した後に、エッジ検出・細線化・ハフ変換を経て検出した直線を図 7 に示す。これより、ほぼ妥当な位置に直線が検出されていることが分かる。さらに、全交点を求めて、その中から実質交点を抽出した結果を図 8 に示す。この図 8 の中で、黒い点は擬似交点、灰色の点は実質交点を示している。さらに、黒い線は位置関係が分かるように便宜上示した、細線化されたエッジである。

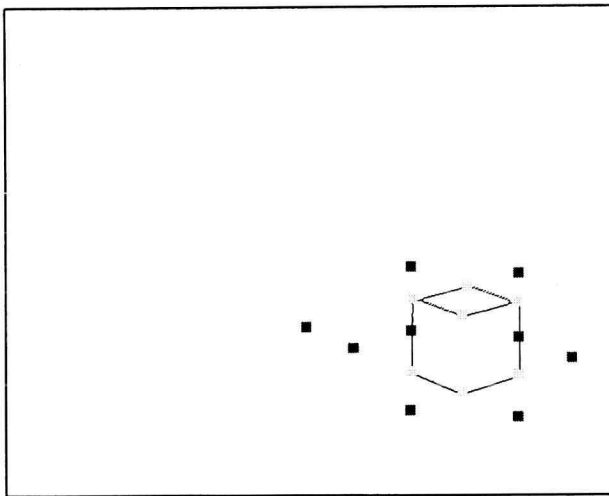


(a) ハフ変換によって検出した直線群 (左画像)

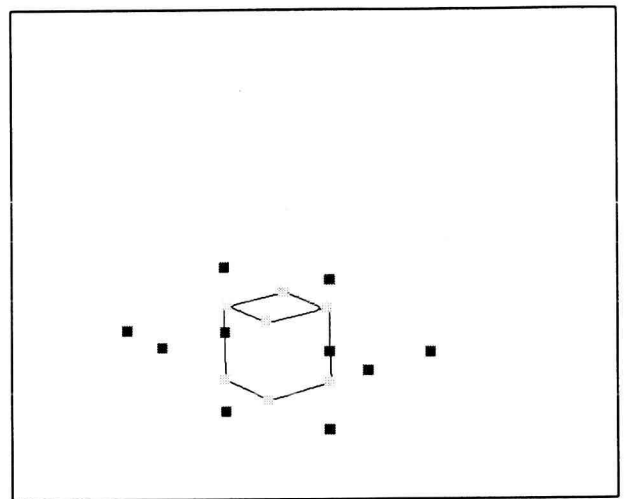


(b) ハフ変換によって検出した直線群 (右画像)

図7 直線の検出結果



(a) 抽出した実質交点 (左画像)



(b) 抽出した実質交点 (右画像)

図8 実質交点の抽出結果

図8に示した実質交点の対応付けを行い、ステレオ法により算出した頂点の三次元座標を表1に示す。なお、表中の頂点番号とは、図9に示すように各頂点に対して割り当てた番号である。

表1 推定した頂点の三次元座標 (単位は[mm])

頂点番号	X	Y	Z
1	33	-18	983
2	4	-24	966
3	56	-25	956
4	25	-31	919
5	3	-60	976
6	57	-61	958
7	26	-67	931

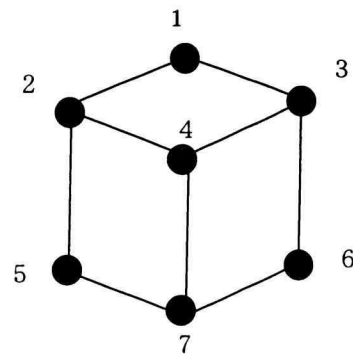


図9 頂点番号

表1に示された三次元座標について、頂点番号が2と5、3と6、4と7の各組でX座標の値がほぼ同じである。このことから、各頂点の相対的な位置関係が妥当であることがわかる。次に、稜線の長さを求めた結果を表2に示す。これについては、真値が45[mm]であることが分かっているので、誤差も算出した。この表から分かるように、絶対値が最大の誤差は-11[mm]であり、真値45[mm]の25[%]程度であった。

表2 算出した稜線の長さとの推定誤差

稜線の両端の 頂点番号	稜線の長さ[mm]	誤差[mm]
1-2	34	-11
1-3	36	-9
2-4	52	7
3-4	49	4
2-5	37	-8
4-7	38	-7
3-6	36	-9
5-7	51	6
6-7	42	3

5. まとめ

本研究の目的は、アクティブビジョンを利用した荷物仕分けロボットの処理・制御アルゴリズムの開発である。これに対して、今回は、カメラを微動して得られる2枚の実画像から、物体の頂点の三次元座標を推定する手法の検討を行った。処理の大まかな流れとしては、エッジ検出・ハフ変換によって検出された全直線相互の交点を求め、その中から実質交点を抽出し、画像間に対応付けを行ってステレオ法を適用するというものである。推定する直線及びその交点には推定誤差が予想されるため、随所で閾値を設けることによりその影響の低減を図った。

実際に、実画像から一辺の長さが45[mm]である立方体の頂点の三次元座標を推定し、稜線の長さで精度の評価を行った。推定誤差は、完全に無視できるほど小さな値ではなかったことから、今後の課題として推定手法の改良が必要である。

参考文献

- [1] 井上哲理, 飯村雅彦, 永森和博, 早川源太, 関靖夫, “物体仕分け操作のための認識の第1次実験的検討”, 神奈川工科大学研究報告 B 理工学編, 第23号(1999).
- [2] 増田大樹, “アクティブビジョンにおける箱型物体の認識の検討”, 神奈川工科大学情報工学科平成15年度卒論(2004).