

アクティブカメラシステムを用いた 3 次元形状の獲得

元 鍾 勳[†] 諸岡 健一^{††} 長橋 宏^{††}

従来のステレオ視はカメラパラメータを固定し、3次元計測を行う研究が多い。本研究ではシーン全体の距離計測を行った後、シーンの中の物体ごとに計測を行い、シーンと各物体との位置関係や物体間の位置関係を確立することにより実空間に近い3次元情報を獲得することを研究の目的とする。物体ごとの計測を行う際、カメラの向きや解像度を調節しながら必要に応じて注目したい物体を計測することにより能動的に3次元計測を行い、計測時間の短縮とデータ量の削減をはかる。

Acquisition of three-dimensional shape with an active camera system

Jonghoon Won[†] Kenichi Morooka^{††} Hiroshi Nagahashi^{††}

There have been a number of studies, which were performed for 3-dimensional measurement by means of fixed camera parameters. In this paper, we will perform 3-dimensional distance measurement in the whole scene, and then measure distance of each object in the scene. It is our aim to establish the position relation between a scene and each object, and the position relation among the objects. When we perform 3-dimensional measurement, we will measure an object by adjusting focus length and resolution to the property of each object. Our methodology makes it possible to perform 3-dimensional measurement actively and to shorten the measurement time, and to reduce the amount of data.

1 はじめに

1.1 研究背景

近年、実空間での3次元距離データの獲得は、仮想世界を現実世界に融合させる「仮想現実」[6][7]、視覚障害者を支援する「障害者支援」[8]、カメラと対象物体間の相対的な位置関係を用いる「自動航海」[9]など、様々な分野で用いられている。また、それ以外にも、認識技術との組み合わせ様々な種類のロボットにも応用されるなど、Computer Visionの分野で、最も重要度の高い課題となっている。

距離計測法は大きく2つに大別される[1]。ひとつは、能動型3次元計測と呼ばれるものである。この方法は、物体にレーザ光や超音波などを投影し、その反射を観測することで物体までの距離を

測定するものである。しかし、カメラだけでなく投影する装置が必要であるため費用がかかるなどの問題がある。また、物体の材質によって投影されたレーザ光に歪みが生じるなど、使用できる物体に制限がある。

もうひとつの距離計測法は、受動型3次元計測法と呼ばれるものであり、その代表的なものにステレオ視がある。ステレオ視とは、平行に設置された2台のカメラから同一対象を撮像し、各画像上の投影位置のずれに基づいてカメラから物体までの距離を計測するシステムである。その画像を取得する際に、人間の視覚と同様に受動的に観測するところに特徴がある。ここでいう受動的とは光を投射したり照明を制御したりと対象物体に能動的に働きかける事はしないという意味である。能動型と比較すると、特別な装置が必要でなく、加えてこの手法は対象物体に制限がないという利点がある。本研究では、ステレオ視を用いて3次元計測を行う。

一般的なステレオ視による3次元計測法は、カメラの向き、焦点距離が固定で、計測領域に制限

[†] 東京工業大学大学院総合理工学研究科
Interdisciplinary Graduate School of Science and Engineering,
Tokyo Institute of Technology

^{††} 東京工業大学大学院理工学研究科
Graduate School of Science and Engineering,
Tokyo Institute of Technology

がある。しかし、実空間のなかには、様々な物体が散乱しており、物体ごとの適確なカメラパラメータはそれぞれ違うものである。

本研究では、カメラの向きや焦点距離を自動的に調節しながら3次元計測を行うアクティブカメラシステムを構築する。

1.2 研究概要

本論文が提案する手法の大まかな流れを述べる。まず、シーン全体の大まかな距離計測を行う。次に、シーン内にある各物体の3次元計測に必要なカメラの向きや焦点距離などを決定する。この決定に基づいてカメラを調整した後、各物体の3次元計測を行う。上述の処理により、各物体間の位置関係を正確に知ることができる。しかし人間は、シーン内にある全ての物体に興味を持つ場合もあれば、1つの物体だけに集中する場合もある。同様に、本研究のシステムにおいても、シーン内にある物体すべてに対して計測を行い、位置関係を確立する必要はない。

本研究の目的は2つである。一つは、カメラの向きと解像度の調節ができるアクティブカメラシステムを構築することである。もう一つは、本システムによる物体の3次元計測法を確立することである。

これらの目的の達成により、ユーザの要求に応じて解像度を調節することで3次元計測精度を向上させることができる。また、必要に応じた物体のみの計測で計測時間の短縮やデータ量の削減ができる。

2 ステレオ視に基づく3次元計測

2.1 ステレオ視の原理

ステレオ視とは、異なる視点から同一の対象を観測し、そのそれぞれの画像上への投影位置の違いから、対象の3次元位置情報を得るものである[2]。

ステレオ視の最も単純な構成は、焦点距離が等しいカメラ2台が、各カメラの光軸が互いに平行で、かつ各々の画像面が同一平面上に載るように配置されている場合である(Fig.1)。

Fig.1 に示すように、ワールド座標系 X 軸は、両カメラの焦点を結ぶ直線に平行な方向とする。また Z 軸は、 X 軸に垂直で且つ画像面と垂直に交

わる方向とする。そして、ワールド座標系の原点は、各カメラの焦点の midpoint である。このようにワールド座標 (X, Y, Z) を定めたとき、空間にある点 $P(X, Y, Z)$ が、左画像内の点 (x_l, y_l) 、および右画像内の点 (x_r, y_r) に投影されたとする。このとき以下の関係式[2][4]が成立する。

$$X = \frac{b(x_l + x_r)}{2d} \quad (2.1)$$

$$Y = \frac{b(y_l + y_r)}{2d} \quad (2.2)$$

$$Z = \frac{bf}{d} \quad (2.3)$$

但し、 f は焦点距離であり、 b は基線長である。また d は視差と呼ばれ、 $d = x_l - x_r$ である。関係式のパラメータの中で、 f と b は既知である。したがって、点 P の左右の投影座標のずれ、つまり視差を求めることで、3次元座標値 (X, Y, Z) が計算できる。

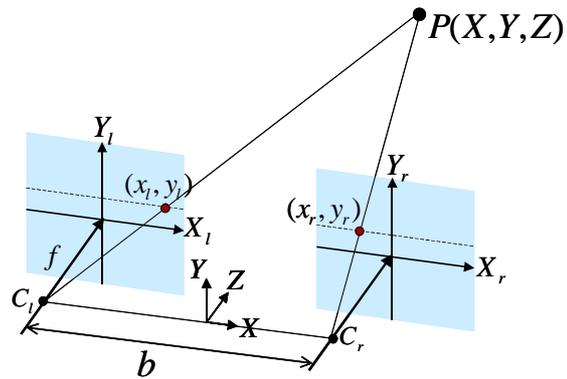


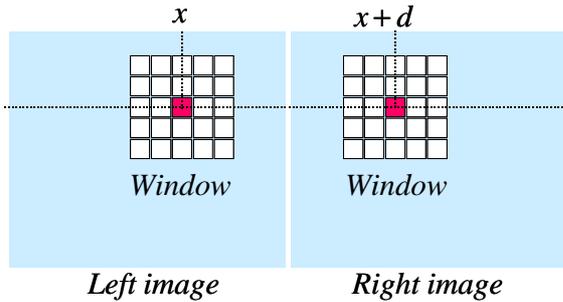
Fig.1 平行ステレオ視

2.2 対応点の探索

ステレオ視の原理では、一方の画像上に投影された各点に対し、その他方の画像からその対応点を求めなければならない。これはステレオ視における対応付け問題[3]と呼ばれるものである。しかし、シーンによっては似たようなパターンが随所に存在することがある。また、異なる視点からの投影による幾何学的歪み、カメラ特性の違い、ランダムノイズ、オクルージョンなどが生じる。これらの原因により完全な対応付けは非常に難しい。

この対応点探索問題はコンピュータビジョンにおける主要な研究分野として長年研究が行われて

おり，この問題の解決法が多数提案されている [10]．これらの手法は主に2つに大別できる．ひとつは，注目点の周囲にウィンドウを設定し，そのウィンドウ内の局所的なパターンから対応点を探索するブロックマッチング法[5]である．もうひとつは，画像上からエッジなどの特徴を抽出し，両画像の特徴間に対応付けを行う特徴法[4]である．本研究では前者を用いて対応点探索を行う．



d : 視差 x : 点の左画像での x 軸座標

Fig.2 ブロックマッチング

Fig.2 に示すように，注目点の周囲にウィンドウを設定し，それをテンプレートとして他方の画像の各画素との一致度を調べる．一致度の算出法としては，ウィンドウ内画素の輝度値の差分の2乗和(SSD)を評価に用いる．具体的な処理は以下の通りである．

$$SSD = \sum_{i=-n}^n \sum_{j=-m}^m \{Y_L(x_l + i, y_l + j) - Y_R(x_r + i + d, y_r + j)\}^2 \quad (2.4)$$

左画像上の点 (x_l, y_l) における輝度値を $Y_L(x_l, y_l)$ ，右画像上の点 (x_r, y_r) における輝度値を $Y_R(x_r, y_r)$ ．ウィンドウサイズを $(2n+1) \times (2m+1)$ とすると，SSDは次のように計算される．そして，この SSD の値が最小となる画素を対応画素とみなし，視差 d を求める．

ブロックマッチングにおける対応点探索では，画像内の全画素を用いると多くの探索時間を要する．この問題を解決するために，エピポラ幾何を用いて探索範囲を限定する．

対応する各画像内の点と計測したい点の3点がある面を形成する．これをエピポラ面と呼ぶ．そして，この面と各画像面との交差によって得られる線をエピポラ線と呼ぶ．この時，一方の画像の点は，もう一方の画像のエピポラ線上にあるという特徴がある．平行ステレオ視の場合，エ

ピポラ線は画像内の同じ高さの画素列となる．したがって，対応点探索は同じ高さの画素列のみで行えばよい．

3 アクティブカメラシステム

3.1 システムの構成

カメラはSONY製のEVI-D30を2台使用する．このカメラは，ズーム，焦点距離，パン/チルトのそとから制御するのが可能である．次に，カメラの設置状況について述べる．各カメラは，パン/チルトのどちらの角度も0度の状態を初期状態と呼ぶ．2台のカメラは，同じ高さで且つ，初期状態のとき焦点の位置が同一直線上になるよう配置する．カメラ間の距離(基線長: Base line)は，今回150mmとした．

本研究のカメラの制御法は，以下の考えに基づく．一方のカメラの向きを大きく変えることによって，両カメラの視線の差が大きくなる．これによって，撮像される共通部分が少なくなる問題がある．この状況を避けるために，2台のカメラのパンの回転角度は常に同じ大きさとする．

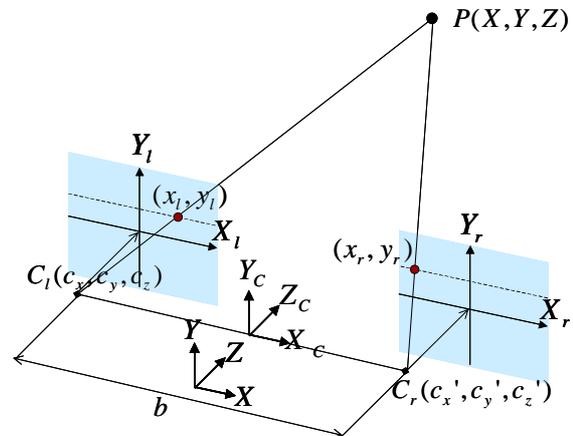


Fig.3 システムの座標系

Fig.3 は，本研究でのワールド座標系とカメラ座標系を示している．上述のカメラの設置状況下で，ワールド座標系の X 軸は，両カメラの焦点を通る直線に平行で，且つカメラの背面を通るよう設定する．ワールド座標系の Z 軸は，カメラの初期状態においてその光軸と平行な方向に設定する．原点はカメラの背面上の中心で，カメラの焦点と同じ高さのところに置く．カメラ座標系の各軸はワールド座標系の各軸と平行で，原点はカメラの焦点を結んだ直線上の中心のところに置く．

画像面の座標系の X, Y 軸は、それぞれ画像の横および縦方向に平行である。また、各カメラの光軸と撮像面との交差点を座標の原点とする。

3.2 カメラパラメータの決定

注目したい物体を計測するためのカメラの向きや解像度の決定法について述べる。

1) カメラの向き

画像面上にある各物体領域の重心を物体の特徴点 $T(t_x, t_y, t_z)$ とする。まず、 T が各画像の中心になるように、両カメラをカメラ座標系の Y 軸周りに式(3.2)で与えられる回転角 σ だけ回転させる。

$$\tan \sigma = \frac{t_x}{t_z} \quad (3.2)$$

次に、回転後の左カメラの焦点 $C_l'(c_x', c_y', c_z')$ からパン角 θ_p 、チルト角 θ_T をそれぞれ式(3.3)、(3.4)を用いて決める。

$$\tan \theta_p = \frac{b}{2d_{OT}} \quad (3.3)$$

$$\tan \theta_T = \frac{t_y}{d_{CT}} \quad (3.4)$$

但し、 d_{OT} は $X-Z$ 平面上でのカメラ座標の原点と T 間の距離、 b は基線長、 d_{CT} は $X-Z$ 平面上での移動後のカメラ C_l' と $T(t_x, t_z)$ 間の距離である。

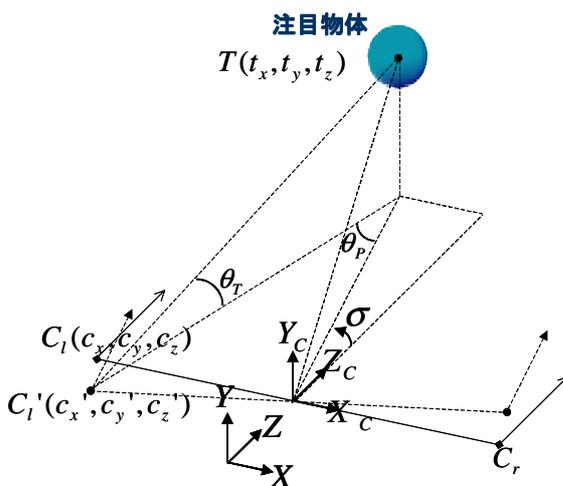


Fig.4 カメラパラメータの決定

2) カメラの倍率

ズームの倍率 N は次のように求める。注目した

物体は、できるだけ歪みの少ない画像の中央部分内で撮像することが望ましい。本研究では、この画像中央領域は、式(3.5)により得られる大きさ S の正方形領域とし、その領域の重心が画像中心と一致するように撮像する。

$$S = \min(W, H) \quad (3.5)$$

但し、 W は画像の幅、 H は画像の高さを表す。次に、画像内の注目したい領域の大きさを、それを完全に覆う正方形領域の面積と見なす。この領域の1辺の長さ s は次式

$$s = \max(w, h) \quad (3.6)$$

を用いて決める。但し、 w は注目物体の画像上の幅、 h は高さを示す。このように、幅と高さを比較して画像面上の物体の大きさを決めることで、シーン内に含まれる様々な物体に適用できる。

式(3.5)および(3.6)を用いて、カメラの倍率 N は

$$N = \frac{S^2}{\lambda s^2} \quad (3.7)$$

によって求める。ここで、 λ は注目物体のズーム操作後のサイズに制限を置く閾値である。今回は $\lambda = 4$ とした。

3.3 対応点探索

平行ステレオ視の場合、左画像の点の対応点は、右画像中のその点と同じ高さの画素列を探索していた。しかし、カメラを回転させた場合、エピポーラ線が変化するため上述の関係が成り立たない。そこで、本システムにおける対応点探索法を提案する。

2 台のカメラを同じ大きさの角度で回転させたとき、エピポーラ線は、

$$y_r = \left(\frac{x_r}{x_r'}\right)\left(\frac{x_l'}{x_l}\right)y_l \quad (3.8)$$

となる。本研究では、式(3.8)を満たす画素を使って対応点探索を行う。

3.4 アクティブステレオ視における3次元位置の算出

左画像に含まれる各画素について対応点を求めた後、注目している点の3次元位置を求める。しかし、2.1 節で示した式(2.1)~(2.3)は、平行ステ

レオ視を仮定して導出した式であるため、我々の手法には適用できない。そこで、本システムにおける3次元位置を求めるための式を新たに導出する。

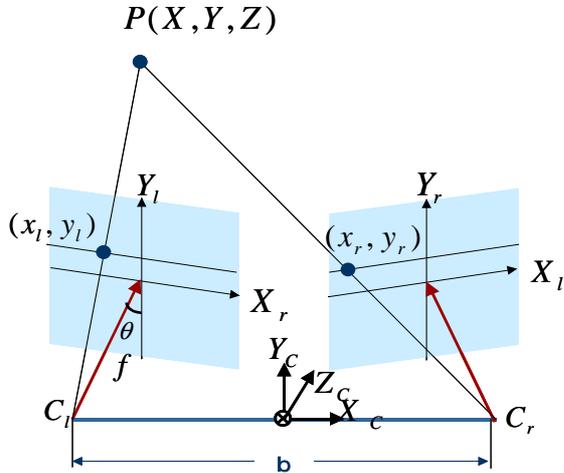


Fig.5 可変ステレオ視の原理

Fig.5 が示すように、ワールド座標系の点 $P(X, Y, Z)$ が左画像と右画像にそれぞれ (x_l, y_l) 、および (x_r, y_r) に投影されているとする。この時、カメラの焦点距離 f 、基線長 b 、パン角 θ を用いて、 (X, Y, Z) はそれぞれ以下の式で得られる。但し、焦点距離はカメラのキャリブレーションにより、あらかじめ求めて置く。

$$X = \frac{b(x_l' + x_r')}{2d'} \quad (3.9)$$

$$Y = \frac{Zy_l}{f \cos \theta - x_l \sin \theta} \quad (3.10)$$

$$Z = \frac{bf \cos \theta}{d'} \quad (3.11)$$

但し、

$$x_l' = \frac{x_l f}{f \cos \theta - x_l \sin \theta} \quad (3.12)$$

$$x_r' = \frac{x_r f}{f \cos \theta + x_r \sin \theta} \quad (3.13)$$

$$d' = 2f \sin \theta + x_l' - x_r' \quad (3.14)$$

3.5 ズームの操作による焦点距離の決定

焦点距離は最も重要なカメラのパラメータで、3次元物体の認識、モデルの再構築などを行う際

にも、カメラと対象物体間の位置関係からその調節が必要であるなど、様々なアプローチから盛んに研究されている[12][13]。本研究では、解像度の調節による計測精度の向上と効率化を図る。

一般的なズームカメラはズームの操作により焦点距離が変わらない特性がある。しかし、式(3.11)から明らかなように、ズームの操作から得た画像を用いて3次元計測を行う時、焦点距離が変わらなければならない。なぜならば、ズームの操作前と操作後の両画像の視差だけ変わって、同じ物体に対し、違う計測結果を得ることになる。このような問題を防ぐために、ズームを変更した時、対象の3次元位置を測るために使用する等価的な焦点距離を求める。本研究ではこの焦点距離を等価焦点距離 f_E と定義し、次にその求め方を説明する。まず、座標が既知である点が描かれたボードをカメラの前に配置し、そこで固定する。そして、ボード内の各ドットの視差を求める。式(3.11)のうち、基線長 b と視差 d とカメラのふれ角 θ とワールド座標の Z 値が既知であることから f_E を導くことができる。このようにして、各ズームに対する f_E を求めるキャリブレーションを行った。キャリブレーションに基づいた f_E の使用により、解像度、および計測精度の向上が可能になる。

4 実験

本論文で提案した手法の有効性を検証するための実験を行った。今回の実験では、物体抽出を容易にするために、対象物体として大部分が白色の人形を、また背景に黒色のカーテンを用いた。そして、物体色を用いて、注目する物体を決めた。

4.1 カメラ向きの調節

人形をワールド座標の原点から Z 軸方向に1000mm離れたところに配置した後、各カメラの中央部にその人形が撮像されるようカメラの向きを調節する実験を行った。まず、平行ステレオ視で左右カメラからの画像を取得した。

Fig.6(a), (b)は、それぞれ左右カメラからの原画像である。次に、両画像の対応点探索を行い、式(2.1)~(2.3)を使って人形の大まかな3次元位置を求めた。Fig.7は視差を濃度値に変換した視差マップである。



(a) 原画像 (左)



(a) 原画像(左)



(b) 原画像 (右)

Fig.6 原画像 (1000mm)



(b) 原画像(右)

Fig.8 回転後の原画像



Fig.7 視差マップ

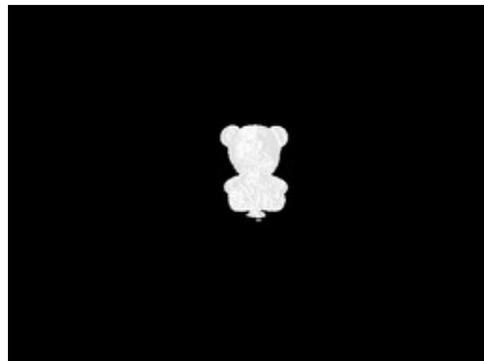


Fig.9 回転後の視差マップ

計測結果から得られた対象物体の重心の座標値は(2,4,1009)である。これを用いて、式(3.3)および(3.4)を使ってカメラの向きに関するパラメータの値をそれぞれ $\theta_p = 4$, $\theta_r = 0$ と求めた。向きの調整を行った後左右カメラから取得した原画像をそれぞれ Fig.8(a), (b)に示す。

Fig.6と比較すると、画像の中央部に物体が撮像されていることが分かる。再び対応点探索を行い、式(3.9)~(3.11)を使って距離計測を行った。Fig.9はその計測結果である。

平行ステレオ視の場合、視差が大きいほど探索領域は広がる。対象物体とカメラ間の距離が短いほど、画像内の多くの点の視差が大きくなり、計測時間が長くなる。カメラ回転前の原画像 Fig.6と回転後の原画像 Fig.8から見られるように、カメラを回転させた場合、左右画像の視差が小さくなる。そのため、探索範囲が狭くなり、その結果対応点探索に要する時間が短くなる。実際、Pentium1Ghzのデスクトップマシン上で、距離計測を行った時間を図った結果を次に示す。

カメラ固定	カメラ回転
1分35秒	27秒

4.2 ズームの調節

次に、画像の中央領域に物体が撮像されるようにズームの調節を行った。式(3.7)を用いて得られた倍率は $N=4$ であった。これにより、ズームの操作を行った。その時に得られた左右の原画像の結果を Fig.10 に示す。



(a) 原画像(左)



(b) 原画像(右)

Fig.10 ズーム操作後の原画像



Fig.11 ズーム操作後の視差マップ

Fig.8 に比べると、より大きく物体を撮像できていることが分かる。画像を背景領域と物体領域と分けた時、Fig.8 では物体領域のピクセルの数が 9894 であり、Fig.10 では、物体領域のピクセルの数が 33939 であった。ズームの操作により、カメラの解像度が上がったことが分かる。提案手法により距離計測を行い、Fig.11 に示す結果が得られた。ワールド座標の原点から Z 軸方向に 1000mm 離れたところに体表点を決め、ズーム操作前後の距離計測を行った結果、それぞれ 1089(mm)、983(mm)であった。ズームの操作により計測精度が上がったことが分かる。

今回の実験では、背景を黒色にして、RGB の値に閾値を置き、背景は計測しないようにした。

Fig.7 と Fig.9 に示した結果で、画像の下部にも計測が行われている。それは背景が光の反射により、閾値内の RGB 値が与えられた結果である。

5 おわりに

シーンの中に存在する物体と、計測システムとの位置関係は様々である。このような状況下でシーンの 3 次元計測を行う際、各物体に対する最適なカメラパラメータは各物体の特徴やカメラまでの距離などによって違う。本研究では、物体ごとにカメラパラメータを調節しながら 3 次元計測を行い、要求される精度を持つ 3 次元データ空間を作るシステムを構築し、それをを用いた 3 次元計測法を提案した。

また、本システムにおける 3 次元位置を求めるための式を新たに導出し、対応点を探索するためのエピポーラ線の式も導出した。

実験を行い、カメラ向きの調節により、探索領域が狭まることで探索時間が約 70% 短縮した。それから、ズームの操作により、解像度が約 3.4 倍向上し、計測誤差が約 80% 減少した。

今回構築したシステムは、単一物体に対するカメラの向きとズームの操作が可能である。しかし、カメラ向きの調節だけでは、ワールド座標の Z 軸上の物体しか注目できない。様々な物体が散在する実空間のなかで、注目可能な領域を広げるためには、カメラ座標の Z 軸の向きの調節と各物体に対するシーンの領域分割が必要である。今後の課題としてカメラ座標の向きも変えるシステムを構

築することが挙げられる。またカメラパラメータの調節に対し、より頑強な対応点探索法を求めることがあげられる。シーンの領域分割は物体の色情報を用いる予定で、現在実験中である。

参考文献

- [1] 井口 征士, 佐藤 宏介, “三次元画像計測”, 昭晃堂.
- [2] 徐 剛, 辻 三郎, “3次元ビジョン”, 共立出版.
- [3] 松山 隆司, 久野 義徳, 井宮 淳, “コンピュータビジョン - 技術評論と将来展望 -”, 新技術コミュニケーションズ.
- [4] 西川 敦, 小川 晋平, 丸 典明, 宮崎 文夫, “アクティブなステレオ視からの隠れ情報に基づく面構造の復元”, 電子情報通信学会論文誌 D-II, vol. J79-D-II, no.2, pp.153 - 164, 1996年
- [5] T.Kanade and M.Okutomi, “A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment”, IEEE Transactions on Pattern Recognition and Machine Intelligence, vol. 16, no. 9, pp.920 - 932, 1994.
- [6] 佐藤 清秀, 山本 裕之, 田村 秀行, “現実空間と仮想空間の位置合わせ手法 ステレオカメラと3次元センサの組合せ”, 画像の認識・理解シンポジウム(MIRU'98)論文集, IPSJ Symposium Series Vol.98, No. 10, pp. 7 - 12
- [7] 佐藤いまり, 佐藤 洋一, 池内 克史, “全方位ステレオによる実光源環境の計測とそれに基づく仮想物体の実画像への重ね込み”, 電子情報通信学会論文誌 D- Vol.J81-D- No. 5 pp. 861 - 871 1998年5月
- [8] Yoshihiro KAWAI, Makoto KOBAYASHI, Hiroki MINAGAWA, Masahiro MIYAKAWA, and Fumiki TOMITA, “A Support for Visually Impaired Persons Using Three-Dimensional Virtual Sound”, Proc. International Conference on Computers Helping People with Special Needs, ICCHP 2000, pp.327 - 334, Karlsruhe, Germany, 2000.7
- [9] 佐藤 淳, 木下 敬介, ロベルト チポラ, “曲面上のエピポラ幾何と未校正カメラによる視覚誘導”, 電子情報通信学会論文誌 D- Vol.J82-D- No. 9 pp. 1401 - 1410 1999年9月

[10] Daniel Scharstein, Richard Szeliski “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms” In IEEE Workshop on Stereo and Multi-Baseline Vision (in conjunction with IEEE CVPR 2001) , pp 131-140, Kauai, Hawaii, December 2001

[11] Reg G. Willson, Steven A. Shafer “What is the Center of the Image?”, Journal of the Optical Society of America A vol.11, No11, 1994, pp2946-2955

[12] W. Brent Seales, “Measuring Time-To-Contact Using Active Camera Control”, CAIP 1995, pp944-949

[13] Peter Lehel, Elsayed E. Hemayed, Aly A. Farag, “Sensor Planning for a Trinocular Active Vision System”, IEEE International Conference on Computer Vision and Pattern Recognition(CVPR'99), Fort Collins Colorado. pp306-312