

## 装着型能動視覚センサを用いた 3 次元注視点の抽出

戸田 真人, 松山 隆司

京都大学大学院情報学研究科知能情報学専攻

**概要** 人と機械が協調して動作する環境を実現するには、システムが人の意図・興味を共有することが必要である。人は、意識を向けた対象を、その時の意図・興味によって異なった形で認識する。そのため、システムにはその対象の位置・形だけではなく、人の意図・興味を含んだ認識をすることが必要である。本研究では、意図・興味を表す情報として人の視線情報を用い、3次元注視点と注視対象の抽出を、視線測定装置とアクティブステレオカメラからなる装着型能動視覚センサを用いて行なう手法を提案する。また、実験により広範囲に渡り、高い解像度で3次元注視点及び注視対象を抽出できることを示した。

## 3D Gazing Point Detection by an Active Wearable Vision System

Masato TODA, Takashi MATSUYAMA

Department of Intelligence Science and Technology,  
Graduate School of Informatics, Kyoto Univ.

**abstract** In order to realize an environment where a machine interacts with human, a system should understand human intention or interest. Human impression about an object depends on his intention or interest. Therefore, the system needs to recognize the object including not only its shape or position but his intention or interest. This paper proposes a method to measure 3D gazing points and an object using an active wearable system with stereo cameras and the Eye Mark Recorder, and demonstrates its effectiveness with experimental results.

### 1 序論

人は外界に対して何らかの行為を行なう時、その対象となる物体に対して何らかの意図・興味を持って関わり合いを持とうとする。したがって、その対象を認識することはその人の意図・興味を理解するための重要な要素となると考えられる。

従来のコンピュータ・ビジョンの分野において、その物体がどこにあってどのような形をしているかなどの“客観的对象認識”が行なわれる。しかしながら、人と共生する人工システムのためのビジョンは、対象の位置・形などの客観的要素だけではなく、人の意図・興味を表す主観的要素を含んだ対象認識が必要となる。本研究では、このような対象認識を行なうための、“主観的对象モデルの構築”を目指している。

本研究では、人の意図・興味を示す情報として、人の視線情報を用いる。人は、動作対象や自分の周囲にあるものなど、様々なものに視線を向けている。

そして、有用であると判断された対象上では、視線は一定時間固視を持続し、対象を“注視”している。注視対象とその3次元位置、および、注視点の対象中の相対位置の移動履歴及び固視継続時間は、その人の行為の意図や、持っている興味、あるいは注意を強く反映していると考えられる。

そこで、その人が注視している対象の主観的对象認識を行なうために、以下の二つのステップに分けて対象のモデル化を試みる。

1. 人の注視点の計測、および、それに基づく注視対象の抽出と記述、また視線の対象に対する相対的な位置や動きの計測を行なうステップ
2. 計測されたデータをもとに、対象の物体を記述し、対象に対する意図・感情を認識するステップ

本研究では図1の視線検出装置と2台のアクティブステレオカメラからなる装着型視覚センサを用いて、研究の第一段階として1の計測ステップのう

ち、注視点抽出方法及び注視対象抽出方法に焦点を当て、これら抽出方法を提案し、実験によりその有用性を示す。



図 1: 装着型視覚センサ

これまで、注視点抽出方法がいくつか提案されている [1, 2]。本研究では、以下の点に問題を置き、注視点抽出方法を提案する。一般にステレオカメラを用いた注視対象を 3 次元的にとらえる方法として形状復元が行なわれるが、対象の位置が離れば離れるほど距離と位置測定精度が悪くなる。高い精度を得るためには高い空間解像度が必要である。一方カメラには人間の視野に合わせて広い視野を持つことが要求される。広い視野を持つためには、全方位ミラーを用いるなどの方法が考えられる [3]。しかし、これらの方法だと視野は広がるが、限られた CCD 面の大きさから解像度が低くなるという問題がある。本研究では、人間が注視している時は限られた範囲しか意識していないことに注目し、パン・チルト・ズームを制御可能なカメラによりステレオ視することで、広い視野内の注視対象を高解像度でステレオ計測する方法を提案する。

本研究では、パン・チルト操作により、カメラの投影中心の移動が生じない視点固定型パン・チルト・ズームカメラを用いる。

以下に本論文の構成を示す。2 章において、必要な座標系の定義とそれらの座標系を関係つけるキャリブレーションについて述べる。3 章において 3 次元注視点抽出方法および能動カメラ制御法、4 章において、注視対象抽出方法について述べ、5 章において結論を述べる。

## 2 座標系の定義とキャリブレーション

### 2.1 座標系の定義

本システムにおいて、問題を記述するために以下の座標系を定義する (図 2 参照)。

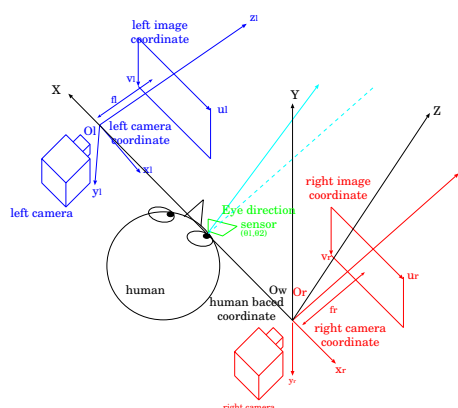


図 2: 座標系の定義

**被測定者座標系** 人の頭部と人が装着したカメラのベースは動かないと仮定し、その共通する座標系を定義する。右カメラの投影中心  $O_r$  を原点  $O_W$  とし、 $X_W$  軸が、 $\vec{O_r O_l}$  と一致するように被測定者座標系  $O_W - X_W - Y_W - Z_W$  を定義する。

**カメラ座標系およびカメラ基準座標系** カメラのパン・チルト動作を含んだ座標系をカメラ座標系として定義する。右カメラの投影中心  $O_r$  を原点とし、 $z_r$  軸がレンズの光軸と一致するように、右カメラ座標系  $O_r - x_r - y_r - z_r$  を定義する。同様に左カメラ座標系  $O_l - x_l - y_l - z_l$  も定義する。

カメラのパン・チルトによって、カメラ座標系と被測定者座標系の関係は変化することになる。カメラのパン・チルトが共に 0 の場合を特別に、カメラ基準座標系として定め、左右それぞれ  $O_r - x_{r0} - y_{r0} - z_{r0}$ ,  $O_l - x_{l0} - y_{l0} - z_{l0}$  を定義する。

**画像座標系** 左右カメラが撮影する画像に対して、画像座標系を定義する。右画像座標系  $u_r - v_r$  は、図 2 のように、 $(x_r, y_r)$  平面に平行で、 $z_r$

座標が右カメラの焦点距離  $f_r$  の位置に画像上でのカメラと光軸の交点  $(u_{r0}, v_{r0}, f_r)^T$  をおくものとする．同様に左画像座標系  $u_l - v_l$  を定義する．

視線角度座標系 視線検出部において，右眼球の回転角度に関して，右眼球回旋中心を原点  $O_\theta$  とした，視線角度座標系  $O_\theta - \theta_1 - \theta_2$  を定義する．ただし， $(\theta_1, \theta_2) = (0, 0)$  方向は右カメラ基準座標系における  $z_{r0}$  軸と平行であり， $\theta_0$  は平面  $O_{r0} - x_{r0} - z_{r0}$  と平行な平面上の角であり， $\theta_1$  は平面  $O_{r0} - y_{r0} - z_{r0}$  と平行な平面上の角であるとする．

## 2.2 キャリブレーション

### 2.2.1 カメラキャリブレーション

カメラモデルとして，視点固定型パン・チルト・ズームカメラモデルを用いる．これらのパラメータを求めるためにカメラキャリブレーションを行なった．

まず，和田らの手法 [4] を用いて内部キャリブレーションを行ない，様々なズームにおける右カメラの内部パラメータを求めた．

次に美越の手法 [5] を用いて外部キャリブレーションを行ない，被測定者座標系とカメラ基準座標系の関係を表す外部パラメータつまり回転行列  $\mathbf{R}_{r0}$  および並進ベクトル  $\mathbf{t}_r$  を求めた．

ここで，右カメラ座標系と右カメラ基準座標系の関係は，回転のみの関係でありそのパン方向の回転角  $\theta_{pr}$ ，チルト方向の回転角  $\theta_{tr}$  を用いて，その回転行列  $\mathbf{R}_{ptr}$  は，

$$\begin{aligned} \mathbf{R}_{ptr} &= \mathbf{R}_{tr} \cdot \mathbf{R}_{pr} \\ \mathbf{R}_{tr} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_{tr} & -\sin \theta_{tr} \\ 0 & \sin \theta_{tr} & \cos \theta_{tr} \end{pmatrix} \\ \mathbf{R}_{pr} &= \begin{pmatrix} \cos \theta_{pr} & 0 & -\sin \theta_{pr} \\ 0 & 1 & 0 \\ \sin \theta_{pr} & 0 & \cos \theta_{pr} \end{pmatrix} \end{aligned} \quad (1)$$

で表すことができる．したがって，カメラ座標系と被測定者座標系の関係を表す外部パラメータのうち，回転行列  $\mathbf{R}_r$  は次式で表される．

$$\mathbf{R}_r = \mathbf{R}_{ptr} \mathbf{R}_{r0} \quad (2)$$

同様に左カメラについても各パラメータを求めた．

本論文では内部パラメータを用いて，歪みを除去した画像を扱うこととする．この場合，被測定者座標系座標から歪み除去後の画像座標への変換は以下の式で表せる．

$$\begin{bmatrix} u \\ v \\ f \end{bmatrix} = fN \left( \begin{bmatrix} \mathbf{R} & | & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \right) + \begin{bmatrix} u_0 \\ v_0 \\ 0 \end{bmatrix} \quad (3)$$

ただし，関数  $N$  は，ベクトル  $(x, y, f)^T$  から  $(x\frac{1}{f}, y\frac{1}{f}, 1)^T$  を求める正規化を行うものであり， $\mathbf{R}$  は回転行列， $\mathbf{t}$  は並進ベクトルである．

### 2.2.2 視線測定装置のキャリブレーション

推定された角度を  $(\theta_1, \theta_2)$  方向にある 3 次元シーン中の点  $M$  が，右画像上でどこに投影されるのかを推定しなければならない．

本論文で用いた視線検出装置では，カメラの撮像面に平行な平面上の点と画像上の点を対応させることによってキャリブレーションを行う．本論文では，右カメラ基準座標系において， $z_{r0} = D_c$  の位置にボード（以下キャリブレーションボードと呼ぶことにする）を置き，ボード中の点と右画像上の点を対応づける．

本研究で使用する視線測定装置では， $(\theta_1, \theta_2)$  を出力せずに，計測された  $(\theta_1, \theta_2)$  を示す変数  $(p_1, p_2)$  を直接  $(u_r, v_r, f_r)^T$  に変換している．視線測定装置の計算ルーチンで用いられる変換式は

$$u_r = f_u(p_1, p_2) = a_{u0} + a_{u1}p_1 + a_{u2}p_2 + a_{u3}p_1^2 + a_{u4}p_1p_2 + a_{u5}p_2^2 \quad (4)$$

$$v_r = f_v(p_1, p_2) = a_{v0} + a_{v1}p_1 + a_{v2}p_2 + a_{v3}p_1^2 + a_{v4}p_1p_2 + a_{v5}p_2^2 \quad (5)$$

であり，共二次方程式と呼ばれる．式 4, 5 より，パラメータ  $a_{un}(n = 0, \dots, 5)$ ， $a_{vn}(n = 0, \dots, 5)$  をそれぞれ決定するために， $(p_1, p_2)$ ， $(u_r, v_r, f_r)^T$  の組が最低 6 組必要になる．EMR-8 では，キャリブレーション時に，この組を 9 点測定して与え，最小二乗法により  $a_{un}$ ， $a_{vn}$  を決定している．

キャリブレーション後の視線測定装置からの計算機への入力は，人の視線とキャリブレーションボー

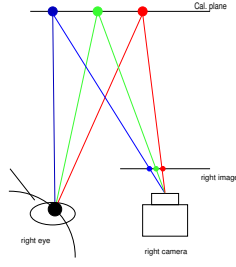


図 3: 視線測定装置のキャリブレーション

ドの交点を右カメラ基準座標系における画像に投影した画像座標である。

### 3 注視点抽出法

#### 3.1 3次元視線直線の導出

被測定者座標系における視線直線の導出法について述べる。

視線測定装置からの出力を  $\mathbf{I}_{Er} = (u_{Er}, v_{Er}, f_r)^T$  とすると、右カメラ基準座標系において視線直線とキャリブレーションボードの交点の位置  $\mathbf{C}_{Er}$  は、

$$\mathbf{C}_{Er} = \left( \frac{u_{Er} - u_0}{f_r} D_c, \frac{v_{Er} - v_0}{f_r} D_c, D_c \right)^T \quad (6)$$

となる。この点は、被測定者基準座標系において次式で表すことができる。

$$\mathbf{C}_{EW} = \mathbf{R}_{r0}^{-1} (\mathbf{C}_{Er} - \mathbf{t}_r) \quad (7)$$

これと被測定者座標系における右眼球回旋中心の位置  $\mathbf{E}_W = (X_{EW}, Y_{EW}, Z_{EW})^T$  を用いて、視線直線  $\mathbf{E}_{view}$  はベクトル方程式を用いて次式で表すことができる。

$$\mathbf{E}_{view} = \mathbf{E}_W + \omega(\mathbf{C}_{EW} - \mathbf{E}_W) \quad (8)$$

#### 3.2 視線直線の画像への投影

3.1 で求められた視線直線を左右画像に投影する。空間中の直線は画像上に投影されても直線になる。したがって、この投影は、視線ベクトル上の任意の2点を画像に投影し、画像上で直線を結ぶことで行うことができる。

被測定者座標系における視線直線上の点  $\mathbf{E}_{W1}$ 、 $\mathbf{E}_{W2}$  を右画像に投影する。それぞれ投影された点

を  $\mathbf{E}_{r1} = (u_{r1}, v_{r1}, f_r)^T$ 、 $\mathbf{E}_{r2} = (u_{r2}, v_{r2}, f_r)^T$  とすると、それらは式3を用いて求められる。同様にして左画像に投影された  $\mathbf{E}_{l1} = (u_{l1}, v_{l1}, f_l)^T$ 、 $\mathbf{E}_{l2} = (u_{l2}, v_{l2}, f_l)^T$  を求める。

したがって、左右画像に投影された視線直線はそれぞれ次式で表すことができる。

$$\begin{aligned} v_r &= a_r u_r + b_r \\ v_l &= a_l u_l + b_l \end{aligned} \quad (9)$$

ただし、

$$\begin{aligned} a_r &= \frac{v_{r1} - v_{r2}}{u_{r1} - u_{r2}}, \quad b_r = \frac{u_{r1}v_{r2} - u_{r2}v_{r1}}{u_{r1} - u_{r2}} \\ a_l &= \frac{v_{l1} - v_{l2}}{u_{l1} - u_{l2}}, \quad b_l = \frac{u_{l1}v_{l2} - u_{l2}v_{l1}}{u_{l1} - u_{l2}} \end{aligned} \quad (10)$$

である。以下、これらの直線をアイエピポーラ線と呼ぶこととする。

#### 3.3 ステレオ画像を用いた注視点探索

注視点を抽出する際、3次元空間中の視線ベクトル上を探索することと画像中のアイエピポーラ線上を探索することは等価である。なぜなら、アイエピポーラ線上の任意の点は3次元空間中の視線ベクトルの一点に対応するからである。また、注視点は十分な大きさをもつ注視対象中にあるため、注視点に対応する左右画像中の点付近においては、同じものが写り、注視点でないアイエピポーラ線上の対応する点はその点周辺部分で異なるものが写る(図4)。したがって、左右画像のアイエピポーラ線上の対応する点の周りの情報を利用することは注視点を抽出する際に大きく役立つと考えられる。左右画像のアイエピポーラ線上のすべての組に対して、その点の周りの情報を含めて一致度を調べ、最も一致度の高い組を注視点とする。

右画像におけるアイエピポーラ線上の点  $\mathbf{P}_r = (u_r, v_r, f_r)^T$  に対応する左画像におけるアイエピポーラ線上の点  $\mathbf{P}_l$  を求める。被測定者座標系における  $\mathbf{P}_r$  の位置  $\mathbf{P}_W$  は

$$\mathbf{P}_W = \mathbf{R}_r^{-1} (\mathbf{P}_r - \mathbf{t}_r) \quad (11)$$

である。 $\mathbf{P}_r$  に投影されている視線ベクトル上の点の被測定者座標系における三次元位置  $\mathbf{P}_v$  は視線直線と右カメラ投影中心と  $\mathbf{P}_W$  を結んだ直線との交点である。

$$\mathbf{E}_W + \omega(\mathbf{C}_{EW} - \mathbf{E}_W) = \xi \mathbf{P}_W \quad (12)$$





で求めることができる．したがって， $\theta_{pr}$  および  $\theta_{tr}$  は次式によって求めることができる．

$$\theta_{pr} = \arctan\left(\frac{x_r}{z_r}\right) \quad (20)$$

$$\theta_{tr} = -\arctan\left(\frac{y_r}{\sqrt{x_r^2 + z_r^2}}\right) \quad (21)$$

同様にして，左カメラにおけるパン方向の回転角  $\theta_{pl}$  およびチルト方向の回転角  $\theta_{tl}$  を求める．

このような左右カメラの制御をすることでズームをかけても抽出できる領域内に注視点があるため正しく注視点を抽出でき，またズームをかけることにより 3 次元位置を精度よく求めることができる．

### 3.5 実験による評価と考察

3.3 節で述べた注視点抽出方法が有用であることを示すために，注視点抽出実験を行なった．

以下の 3 つの状態での注視点抽出を行なった．

1. カメラ制御を行なわない状態での注視点抽出
2. 視線方向に合わせたパン・チルト制御を行なった上での注視点抽出
3. 注視点の 3 次元位置より，パン・チルト・ズームコントロールを行なった上での注視点抽出

実験結果を図 3.5 に示す．画像上において白い直線がアイエピポーラ線，楕円の点が抽出された注視点である．実験結果より以下のことが言える．

- どの場合も，画像上で同一の点を示している．このことは，カメラの状態にかかわらず注視点を抽出できることを示している．
- カメラ制御を行なわない場合は対象（机にのっている箱）が全て共通視野内に収めることができないが（右画像に写り切らない），カメラのパン・チルト制御を行なうことで十分収めることができる．
- 注視点のおおよその位置情報を元にパン・チルト・ズーム制御を行なうことで，解像度を高めた状態での注視点探索を行なうことができ，より詳細な注視点の 3 次元位置を抽出することができる．

以上のことより，本論文で提案した注視点抽出方法により，広範囲に渡り精度よく 3 次元注視点を抽出できると考えられる．



カメラ制御をしない  
場合 (左画像)



カメラ制御をしない  
場合 (右画像)



パン・チルト制御をした  
場合 (左画像)



パン・チルト制御をした  
場合 (右画像)



パン・チルト・ズーム  
制御をした場合 (右画像)



パン・チルト・ズーム  
制御をした場合 (左画像)

図 5: 能動カメラ制御実験結果

## 4 3次元注視対象抽出

### 4.1 ステレオ画像を用いた点ベースでの 3次元位置測定

テンプレートマッチングを用いて左右画像の対応点を探索し，その対応する点同士の情報を用いてその 3 次元位置を測定する．

テンプレートマッチングの手法 [6] を用いて右画像の点  $m_r = (u_r, v_r, f_r)^T$  と左画像の点  $m_l = (u_l, v_l, f_l)^T$  の対応を捜し出す．この情報よりその点の被測定者座標系における 3 次元位置  $M = (X, Y, Z)^T$  を求める．

### 4.2 注視対象の抽出

距離画像と注視点の位置より注視対象の抽出を行なう．右画像における注視点の位置を  $m_e = (u_e, v_e)$  とする．以下に注視対象の抽出方法を示す．

1. 注視点は、注視対象中の1点である。これを初期値として与える。
2. 注視対象中にあると判断されている画像中の点の距離とその周りの4近傍の点の距離を4.1節で上げた方法で求め、それぞれ比較することにより、その点が同一対象中にあるか判断する。

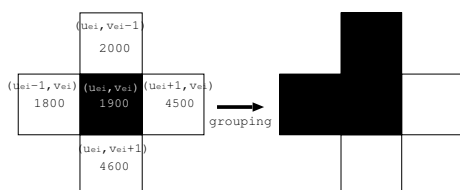


図 6: 注視対象抽出

3. 2を繰り返し行なうことによって、同一対象中にある点をまとめていくことで注視対象を抽出する。

本論文で提案した手法を用いると注視点の3次元位置情報およびステレオ画像より、注視対象の3次元位置情報のみを抽出し、他の部分の3次元位置情報を抽出することがなく、計算量が少なくすむ。

### 4.3 実験による評価と考察

以下に実験方法を示す。

1. 3次元注視点を抽出する。
2. 4.2節の方法より、注視対象の抽出及び注視対象の距離画像を作成する。

本実験ではテンプレートのサイズとして  $32 \times 32$ [pixel<sup>2</sup>], 領域結合の閾値を 30[mm] として実験を行なった。実験結果を、図 4.3 に示す。実験結果より、以下のことがいえる。

- 対象内の領域が対象として、抽出されている。このことは提案した手法に基づいて、対象抽出が行なえることを示している。
- 対象全体ではなく、一部欠けている部分がある。また対象でないものも抽出されている。これは、テンプレートマッチングによる対応づけの問題であると考えられる。

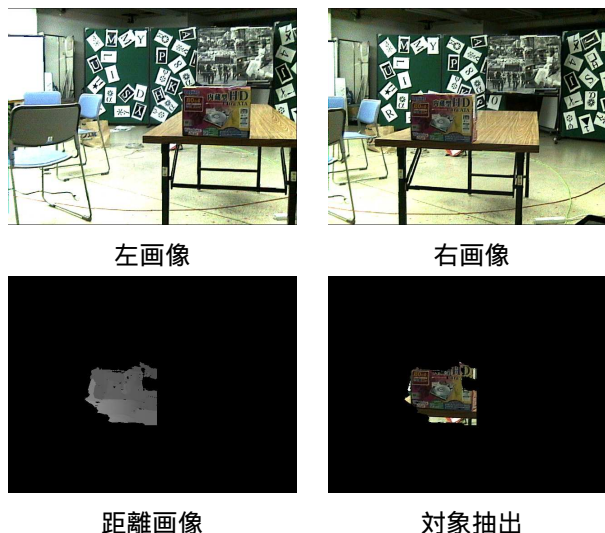


図 7: 注視対象抽出実験

以上のことより、テンプレートマッチングによる対応点探索の問題はあるが、全体の距離画像を出しての抽出に比べ、同じ精度を保ちながらも、より計算量が少なく対象抽出を行なうことがいえると考えられる。

## 5 結論

本論文では、視線情報を用いた主観的对象認識のための計測として、アクティブステレオカメラと視線測定装置を備えた装着型視覚センサを用いた3次元注視点の抽出および3次元注視対象の抽出方法を提案し、実験によりその有用性を示した。

今後の課題として、以下のことが上げられる。

**対応点探索アルゴリズム** 本論文では、注視点を抽出する際に対応する画素の差を評価した。この方法であるとテクスチャによっては正確に抽出できない場合がある。エッジ画像を用いる場合も考えられるが、この場合も視差の影響が強くなり正確に注視点が抽出できない場合がある。したがって、テクスチャや視差の変化に強い、画像を扱う方法を考える必要があると考えられる。

**リアルタイム性の向上** 実際にカメラ制御を行なう時に、カメラの回転させたりズームの変化させるには多少の時間がかかる。本論文では、その間注視点は変化しないものと考えている。しか

し、実際はこの間に注視している場所が変わることが多い。したがって、人間の視線にあわせてパン・チルト・ズーム制御ができることが望ましい。これを解決するためには、人間は距離に対してどのような視線・眼球の動きをするのかなどのこと、つまり人間の生理的特性を調べ、それをシステムに組み込むということをする必要があると考えられる。

主観的对象の特徴化 本研究の目的は、主観的对象認識のための対象のモデル化である。本論文で提案した手法により、注視点の3次元位置、また注視対象の3次元位置を抽出することができた。これらの情報より主観的对象の特徴化をし、モデル構築をする必要がある。

## 謝辞

本研究は科学研究費補助金特定領域研究 13224051 の補助を受けて行った。

## 参考文献

- [1] A.Sugimoto,A.Nakayama and T.Matsuyama:”Detecting a Gazing Region by Visual Direction and Stereo Cameras”,Proc.of ICPR2002 Vol.3,pp.278-282,2002.
- [2] 満上育久, 浮田宗伯, 木戸出正継: 視線情報を用いた注視点の3次元位置推定, 電子情報通信学会技術研究報告 PRMU2002-169.
- [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”, p13-17, 1998.
- [4] 松山隆司, 和田俊和, 物部祐亮: 視点固定型パン・チルト・ズームカメラを用いた実時間対象検出・追跡, 情報処理学会論文誌, Vol.40, No.8,pp.3169-3178, 1999.8.
- [5] 美越剛宣: 平面を利用した3次元画像計測, Master's thesis, 京都大学, 1998.
- [6] 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”, p123-137, 1998.