

協調3次元センシング技術を用いた人物検出手法に関する研究

- 生活活動を見守る知的空間の構築に関する研究 -

渡辺 博己 棚橋 英樹 浅井 博次
平湯 秀和* 清水 早苗* 山本 和彦**

Study on a Method of Moving Objects Detection Using Cooperative 3-D Vision - Construction of Intelligent Environment for Life Supports -

Hiroki WATANABE Hideki TANAHASHI Hirotsugu ASAI
Hidekazu HIRAYU* Sanae SHIMIZU* Kazuhiko YAMAMOTO**

あらまし 治安の悪化や高齢社会の進展に伴い、ユビキタスコンピューティングを中心とした知的空間技術を用いて生活の安全性、快適性、利便性を向上させることが重要な課題となっている。そこで、知的空間を構築する上で重要となる人物追跡技術に着目し、複数の3次元センサを協調させた人物検出手法の検討を行い、その有効性を評価した。本報告では、知的空間を実現するためのプロトタイプシステムを構築し、複数人物が存在する環境下での人物検出を行った。

キーワード 知的空間、全方向ステレオシステム(SOS)、多視点ステレオカメラ、人物検出

1. はじめに

治安の悪化や高齢社会の進展に伴い、個人生活の安全性、快適性、利便性向上を支援するサービスの需要が増加している。また、コンピュータやセンシング・通信機器の高機能化、小型化により、人の周りでさりげなく見守り、日常生活空間のあらゆる場所で生活をサポートしてくれるユビキタスコンピューティング技術への期待と要望が年々高まっている。

そのような中で、家庭にユビキタスコンピューティング技術を取り入れ、家庭環境を知的化することにより、安心、安全、快適な生活環境を実現する知的空間技術の研究が数多く行われている^[1-6]。例えば、米国ジョージア工科大学のAware Homeでは、センサとコンピュータによって家族の状況を把握してくれる家と、その機能を利用したアプリケーションが研究されている^[1]。また、(独)情報通信機構のユビキタスホームでは、家電製品だけでなく様々な機器の相互接続性の実証実験や、機能分散協調基盤を中心とする家全体のロボット化、その中での人

の行動把握とそれに基づく新たなサービスが研究されている^[4]。

様々な生活シーンにおいて、生活に応じたサービスを人に供給するためには、「如何に環境を理解できるか」に集約される環境センシング技術が重要であり、画像センシング技術は解決すべき課題の一つである。画像センシング技術は、人にセンサを装着させることが無いため、日常生活を束縛することなく生活行動を取得できると共に、人とその周りの環境の状況を同時に取得できるため、人と人、人と物とのインタラクションの理解などに有効である。特に、人物追跡技術は人物の動線や行動パターンを取得できるため、不審者の監視だけでなく、居住者の状況を察知することが可能となる。また、3次元空間上で人物を追跡することは、画像上で人物同士が重なり合っている場合でも、精度よく人物を分離することができるため、個々の人物の行動パターンを容易に取得することが可能となる。

そこで本報告では、高精度に人物追跡を行うための人物検出手法について検討し、複数の3次元センサを協調させた人物検出手法について述べ、その有効性を評価する。

* (財)ソフピアジャパン

** 岐阜大学 工学部

2. 人物検出手法の検討

2. 1 多視点カメラを用いた人物検出

多視点カメラを用いた人物追跡の研究として、観測情報をカルマンフィルタにより統合して人物追跡を効率的に行う研究が行われている^[7]。時刻間での人物の対応をとるための特徴には、人物シルエットの中央点を用い、人物追跡には予測される人物位置にガウス分布的な存在確率を用いている。2名程度ではあるが良好な人物追跡を実現している。しかし、シルエット画像において人物のシルエットを分離できない場合には検出が困難になる。

また、岐阜県地域結集型共同研究事業^[10]（以下、結集事業）では、3次元形状復元に用いられる視体積交差法に人物の存在確率マップを用いることにより人物追跡を行った^[11]。複数の異なる視点から得られたシルエット画像を空間に投票し、和となる部分に存在確率マップを適用することで、投票によって生じたノイズボクセルの除去と人物同士の分離問題を解決している。それにより、人物の床面上の位置だけでなく、投票空間中の人物ボクセルから身体各部位の像を検出することができた。しかし、人物追跡の成功率は、人物の空間侵入時における存在確率マップの当てはめに大きく依存するため、空間への人物の侵入に対し、存在確率マップの当てはめを容易にする拘束条件を与えることで高い追跡精度を実現している。

2. 2 ステレオカメラを用いた人物検出

ステレオカメラを用いた研究として、複数の人物を精度よく分離し、追跡を行っている研究がある^[8]。3次元計測した特徴点群を3次元空間上でクラスタリングすることで人物検知を行い、位置情報に基づいて追跡を行っている。また、同一空間を取り囲むように複数のステレオカメラを配置して人物の行動理解を行う研究がある^[9]。個々に得られた距離情報を3次元空間上で統合した後、床面に水平な複数段階の平面に写像することで人物を検出し、姿勢の認識を行っている。しかし、これらは、カメラ画角の制限により、観察対象空間とカメラ設置位置との間に一定の距離を置かなければならない。つまり、観察対象空間を大きくするほど、カメラを対象空間から離して設置する必要がある。

2. 3 多視点ステレオカメラを用いた人物検出

結集事業で開発した全方向ステレオシステム^[12]（以下、SOS）は、画角の制限がなく、設置位置を考慮する必要がないため、センサを対象空間に自由に設置することで、距離情報を用いた人物追跡を行うことが可能である^[16]。しかし、ステレオカメラを用いた研究と同様に、観察対象物体とシステムとの間の距離が大きくなるに従ってステレオ精度が低下する。

そこで、室内空間における人物追跡を実現するために、観測対象空間を取り囲むように複数のSOSを設置することにより、ステレオ計測による距離情報と視体積交差による空間情報を用いた人物検出手法を提案する。具体的



図1 ASTRO-Sensorシリーズの概観
（左から順にVenus，Mercury，Saturn，Jupiter）

には、ステレオ計測で得られた距離情報に信頼度を付加することでステレオ精度の低下による計測誤差を低減し、視体積交差法により得られた人物候補領域に信頼度付き距離情報を用いることでノイズボクセルを除去する。これにより、存在確率マップの適用をスムーズにし、視体積交差法を用いた人物検出手法における拘束条件を緩和した人物検出が可能になると考えられる。

3. プロトタイプシステムの構築

3. 1 埋め込み型SOSの開発

結集事業では、センサを中心とした全方向のカラー画像と距離画像を完全に死角無く、高い解像度でリアルタイムに取得することが可能なSOS^[12]を提案してきた。これに加え、移動体ビジョンを指向し、小型軽量化を図った小型全方向ステレオシステム（miniSOS）^[13]、及びサーベイランスシステム等の用途を指向し、システムサイズを小型化しながらも10cmのステレオベースラインを確保したモデル^[14]、さらに筒状のカメラブロックを持つ1/4球状のシェルを結合する構造により、レンズを保護しつつ剛性を増した高剛性モデル^[15]を開発し、初代モデルをJupiter、移動体ビジョン指向型をVenus、サーベイランスシステム指向型をSaturn、高剛性モデルをMercuryと名付け、ASTRO-Sensorシリーズとして用途に合わせたシリーズ展開を図ってきた（図1）。しかし、家庭のような生活空間で固定して使用する際には、全方向を死角なくセンシングできる性能ゆえに、配置方法を考慮する必要がある。

そこで、従来モデルの視野範囲が全球をカバーしているのに対し、その1/8の視野をカバーする構成とすることで、天井の隅に埋め込むような形で固定できる1/8球型モデルを開発した。図2に1/8球型SOSの外観、表1に緒元を示す。

本センサは、3つの3眼ステレオユニットから構成され、1/8球の視野を持つように設計されている。ステレオユニ

表1 ASTRO-Sensorシリーズと1/8-SOSの緒元

モデル	Jupiter	Saturn	Venus	Mercury	1/8球型
基本形状	正20面体	正12面体		球状	-
撮像素子	1/3 カラーCMOS		1/4 カラーCMOS		
素子の解像度	640(H) × 480(V) pixel				
焦点距離	2.9mm	2.5mm	1.9mm	3.3mm	
画角	96.6 ° (H) × 71.9 ° (V)	115.5 ° (H) × 83.8 ° (V)	101 ° (H) × 76 ° (V)	59.5 ° (H) × 45.4 ° (V)	
ベースライン	90mm	100mm	50mm		
フレームレート	15fps				
カメラヘッド直径	27cm	21cm	11.6cm	12cm	-
重量	4.5Kg	1.85Kg	615g	846g	-

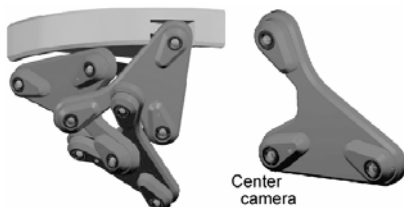


図2 1/8球型SOS (左) とステレオユニット (右) の外観

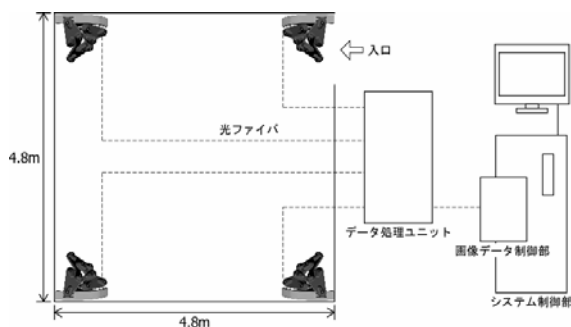


図3 試作したプロトタイプシステムの構成

ット上の3つのカメラのうち、センタカメラはカラー画像を取得し、他の2つのカメラはモノクロ画像を取得する。ステレオユニット上の各カメラは同一平面上にあり、それぞれの光軸は互いに平行である。また、センタカメラからステレオ対を成す2つのカメラまでのベースラインは直交している。これにより、各ステレオ対がそれぞれ水平垂直のエピポーラ拘束を満たし、対応点探索の処理コストを低減することが可能である。

また、1/8球視野を3つのステレオユニットでカバーすることで、正12面体形状を持つVenusやSaturnの倍の解像度が得られる。実際には、VenusやSaturnは各ステレオユニット間での重なり領域が1/8球型モデルよりも大きい。そのため単純に比較することはできないが、角度分解能が向上し、ベースラインがVenusと等しいにもかかわらず、Venusに比べ高いステレオ精度が得られる。これにより、設置位置からの全方向性を残しながらも違和感の少ない配置と、小型ながらも高いステレオ精度の確保が可能となる。

3.2 システムの構成

図3に人物検出のために構築したプロトタイプシステムの構成を示す。本システムでは、空間の1辺が480cmとなるようにパーティションで区切り、その天井の4隅それ

ぞれにセンサを設置した。天井までの高さは270cmである。

各センサで取得される3枚のカラー画像と6枚のモノクロ画像は、光変換機能を有するデータ処理ユニットに転送される。データ処理ユニットは4つのセンサから転送された36枚の画像データを統合し、光ケーブルにより1.2Gbps × 2の光信号として画像データ制御部へ出力する。画像データ制御部は高速シリアル信号として送られてきた画像データを復調し、システム制御部のメモリへ高速に転送する。システム制御部では、ステレオユニット毎にキャリブレーションデータを管理し、ステレオビジョンソフトウェアライブラリ^[18]を用いて、補正画像、距離画像を生成し表示する。なお、センサ間は1つの基準座標系を用いて回転、並進パラメータを求め、センサ上のステレオユニット間には高い精度で組み付け加工された配置パラメータを用いてキャリブレーションを行っている。個々のステレオユニットに関しても、それぞれユニット単位でキャリブレーションを行っており、レンズ歪み及びカメラ間のアライメントずれの影響をソフトウェアによって除去している。

4. 複数人物の検出

室内で人物を追跡するために、環境に埋め込んだ複数の1/8球型SOSを用いて人物を検出する。以下に、処理の流れを示す。

- 1) あらかじめ生成した背景モデルと入力画像とを比較し、差分を求めることでシルエット画像を生成する。
- 2) 得られたシルエット画像に基づいてボクセル空間に投票を行い、人物候補領域を検出する。
- 3) 距離画像から得られた3次元情報に信頼度を付加し、信頼度に基づいて距離情報をボクセル空間へ写像することにより、ノイズ領域を除去し、人物領域のみを抽出する。

ここでは、それぞれの処理について詳細を述べる。

4.1 シルエット画像の生成

シルエット画像は、各ステレオユニットのセンタカメラ画像に対し、Radial Reach Correlation (RRC)^[17]により、それぞれ前景領域を抽出するとともに、距離画像に対しても背景差分を行い、2枚のシルエット画像の論理積を求



図4 生成したシルエット画像の例

めることにより生成する。

単純背景差分を用いる場合，出現物体の明度分布と背景の明度が近い領域において未検出が生じるだけでなく，影などの明度変化による過検出の影響も大きい．一方，RRCは濃淡情報をベースに，ピクセル単位の分解能で局所的なテクスチャを評価し，変化領域の明度分布に依存せず，しかも影などの明度変化の影響を低減したロバストな検出が可能である．また，距離画像における背景差分では，距離が近くなった画素についてのみ検出した．これは，照明等の影響による距離の揺らぎで生じた誤検出を低減するためである．

図4にシルエット画像の生成例を示す．図4(c)はRRCのみで背景差分を行った結果であり，図4(d)は距離画像のみで背景差分を行った結果である．図4(e)は，2つの差分画像の論理積により求められたシルエット画像に対し，膨張・収縮処理を加えたものである．

4.2 視体積交差による投票

得られたシルエット画像とキャリブレーションデータに基づいて，視体積交差により投票を行うことで，ボクセル空間上で人物候補領域の検出が可能である^[11]．

視体積交差法では，カメラ位置とシルエット画像の各画素とを結び直線上でボクセル投票を行う．通常，視体積交差法による形状復元では，精度を高めることが重要であるため，高い空間分解能をもつボクセル空間を用いる必要がある．しかし本稿では，人物検出を目的として視体積交差法を用いているため，形状の精度はあまり重要ではない．そのため，ボクセル空間の分解能を低く設

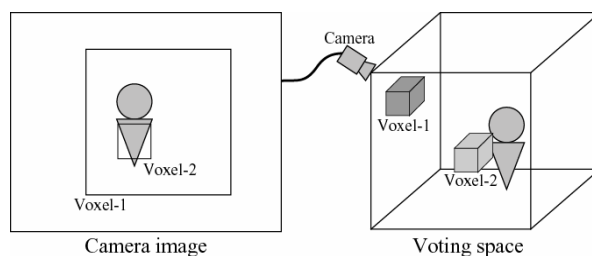


図5 投票空間とカメラ画像との関係

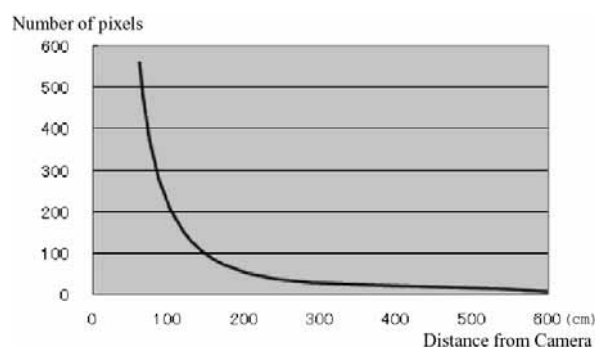


図6 カメラからの距離と画素数との関係

定することが可能である．このような場合，ボクセル空間の分解能に対して，カメラ画像の解像度が高くなるために，カメラに近い位置のボクセルでは，多重に投票されるといった問題が生じる．そのため，投票時には多重投票の回避を行うなどの判定処理が必要となる．図5に投票空間と画像の関係を模式的に示す．

各ボクセルに対応する画像中の領域のサイズは，カメラからボクセルまでの距離によって大きく異なる．つまり，カメラに近い位置のボクセルに対応する範囲は大きく，カメラから遠い位置のボクセルに対応する範囲は小さい．そこで，各ボクセルに対応する画像領域内のシルエットの割合により，ボクセル投票を行うか否かを決定する．位置 (x, y, z) に対するあるカメラ C からの投票度数 $V_C(x, y, z)$ は，画素 F_C の平均値

$$V_C(x, y, z)_{ave} = \frac{1}{K} \sum_{j=1}^K F_C(i, j) \quad \dots\dots\dots (1)$$

$$s.t. \quad F_C(i, j) \in T_V(x, y, z)$$

より，

$$V_C(x, y, z) = \begin{cases} 1 & \text{if } V_C(x, y, z)_{ave} \geq th_1 \\ 0 & \text{otherwise} \end{cases} \quad \dots\dots\dots (2)$$

として求まる．ここで $T_V(x, y, z)$ は，ボクセル $V(x, y, z)$ に対応する画像領域を示し， K は画素数である．図6にカメラからの距離と画素数の関係（ボクセル空間解像度5.0cm）を示す．カメラに近いほど画素数は極端に増すが，画像領域の大きさは異なっても，空間分解能は一定であることから，閾値 th_1 は対象の大きさによって一意に決定

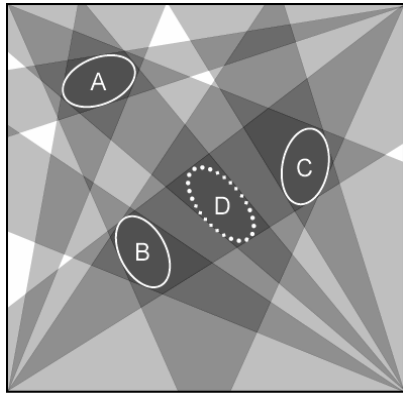


図7 視体積交差による投票の例

できる．また，人物検出を目的とした場合，空間分解能が人物に対して十分小さければ， th_1 は高く設定することができる．なお，画像解像度をカメラから最も遠い位置での投票空間の感度と対応させることで，無投票ボクセルの発生を回避することができる．

ボクセル V への投票度数は， N 台のカメラから，

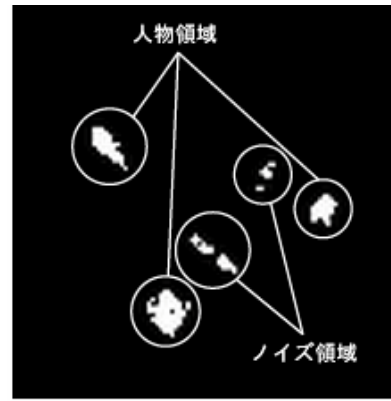
$$V(x, y, z) = \frac{100}{N} \sum_{C=1}^N V_C(x, y, z) \geq th_2 \dots\dots\dots (3)$$

として求められ，閾値 th_2 によって人物候補ボクセルとして判断される．閾値 th_2 はカメラ台数より決定できる．

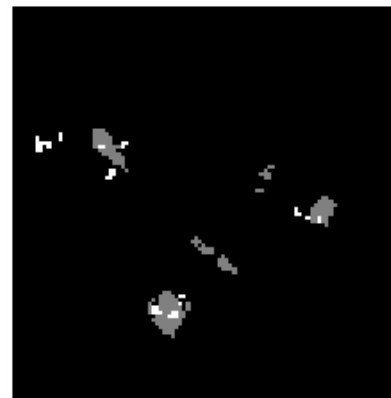
4. 3 人物領域の抽出

視体積交差による投票からは，人物領域だけでなく，人物同士のオクルージョンによるノイズ領域も人物候補領域として取得される．図7に各センサからの投票例を示す．図中実線の楕円領域(A～C)を実際の人物領域とすると，グレーの領域が各シルエット画像からの投影結果として投票される．ここで，グレー領域の濃淡は投票度数を示しており，濃くなるほど投票度数が高い．投票度数の高い領域を人物候補領域として求めると，Dの領域のように人物同士のオクルージョンにより投票度数が高くなる領域が存在する．これを解決する方法として，人物が空間に侵入した際に存在確率分布を当てはめ，人物を追跡することによりノイズ領域を除去する方法がある^[11]．しかし，この方法では，最初のフレームで検出を失敗した場合，その後の人物検出に大きな影響を及ぼす．そこで本報告では，距離画像から得られた3次元情報をボクセル空間へ写像することにより，ノイズ領域を除去し，人物領域のみを抽出する方法を試みる．

センサからはカラー情報と共に距離情報も同時に取得することができるため，各センサより得られた3次元座標を基準座標系に変換し，視体積交差で使ったボクセル空間に写像する．しかしながら，ステレオカメラで得られた距離画像は，カメラと対象との距離が遠くなればなるほど精度が低下する．また，本センサは得られた画像中のテクスチャのみに依存するパッシブステレオ方式を採用しているため，測定に有効なテクスチャの存在しな



(a) 視体積交差により投票されたボクセル



(b) 人物の3次元座標を写像したボクセル

図8 人物領域の検出例

い領域に対しては情報が欠落し，安定性の面でやや問題がある．そこで，本報告では，得られた3次元座標をボクセル空間に写像する際には，3次元座標の各カメラからの距離 $D_C(x, y, z)$ に応じた

$$E_C = \frac{M}{D_C(x, y, z)} \dots\dots\dots (4)$$

より与えられる信頼度 E_C を求め，

$$W(x, y, z) = \sum_{C=1}^N E_C \geq th_3 \dots\dots\dots (5)$$

により， N 台のカメラから得られる信頼度の総和をボクセル空間と同一の空間分解能を持つ3次元座標空間 $W(x, y, z)$ に写像した．ここで， M は定数である．なお，写像する3次元座標は，シルエット画像で得られた画素を用いる場合，ステレオ処理におけるバウンダリオーバーリーチの影響により，人物領域だけでなく背景領域も含まれる場合があるため，シルエット画像を収縮した画像の画素を用いた．

最終的に，人物領域は $W(x, y, z)$ を含む人物候補ボクセル群として求めることができる．図8に視体積交差法により投票されたボクセルと，それに信頼度を付加した3次元座標をボクセル空間に写像した結果を示す．図はボクセル空間を上から地平面に透視した結果であり，図8(b)の

白色の領域が写像された3次元座標である。人物領域に3次元座標が正しく写像され、ノイズ領域を除去することが可能であることを示している。

5. まとめ

全方向ステレオシステムの特性を活かしながら、環境に埋め込むことができる1/8球視野のステレオセンサを開発し、それを複数利用して実現する人物検出手法について述べた。本センサは天井の隅からの全ての方向のカラー画像と距離画像を死角無くリアルタイムに供給することが可能で、全球型の同一ベースラインを持つモデルよりも高い空間分解能を持つ。また、本センサを複数利用することにより、視体積交差法とステレオ法とを利用したハイブリッドな人物検出が可能であることを示した。その結果、簡単な処理でノイズ領域を除去することが可能となった。

構築したプロトタイプシステムでは、3名の人物を対象にノイズ領域の除去と人物領域の抽出を行ったが、今後はさらに多くの人物に対して実験を行う予定である。また、リアルタイムでの人物追跡を目標とし、アルゴリズムの高速化を検討すると共に、追跡情報を利用した生活空間でのアプリケーションを開発していく予定である。

謝 辞

本研究の遂行にあたり、1/8球型全方向ステレオシステムを貸与して頂いた(財)ソフトピアジャパンIT研究センターの丹羽義典センター長をはじめ、関係者の皆様に感謝します。また、有益な助言を頂いた早稲田大学WABOT-HOUSE研究所の富永将史氏に感謝します。

文 献

- [1] Aware Home, <http://www.awarehome.gatech.edu/>
- [2] Easy Living, <http://research.microsoft.com/easyliving/>
- [3] Smart Rooms, <http://vismod.media.mit.edu/vismod/demos/smartroom/>
- [4] コピキタスホーム, http://www2.nict.go.jp/jt/a135/research/ubiquitous_home.html
- [5] 人を見守るデジタルヒューマン, http://www.dh.aist.go.jp/research_new/enabling/index.php
- [6] T. Mori, H. Noguchi, A. Takada, T. Sato, "Sensing Room: Distributed Sensor Environment for Measurement of Human Daily Behavior," First International Workshop on Networked Sensing Systems (INSS2004), pp.40-43, 2004.
- [7] 森大樹, 内海章, 大谷淳, 谷内田正彦, 中津良平, "非同期多視点画像による人物追跡システムの構築", 信学論(D-II), Vol.J84-D-II, No.1, pp.102-110, 2001.
- [8] 池谷崇, 海上尚子, 諏訪正樹, 来海雅俊, "ステレオビジョンセンサを用いた人物検知", 第10回画像センシングシンポジウム(SSII04)講演論文集, pp.231-234, 2004.
- [9] 依田育士, 坂上勝彦, "コピキタスステレオビジョンによる実時間実環境ヒューマンセンシング", 第9回画像センシングシンポジウム(SSII03)講演論文集, pp.397-402, 2003.
- [10] 岐阜県地域結集型共同研究事業, <http://www.softopia.or.jp/rd/index.html>
- [11] 富永将史, 本郷仁志, 奥水大和, 丹羽義典, 山本和彦, "人物行動把握のための複数人物追跡システムの構築", 動的画像処理実利用化ワークショップ(DIA2004), pp.79-84, 2004.
- [12] 山本和彦, 棚橋英樹, 桑島茂純, 丹羽義典, "実環境センシングのための全方向ステレオシステム(SOS)", 電学論, Vol.121-C, No.5, pp.876-881, 2001.
- [13] 佐藤雄隆, 山本和彦, 桑島茂純, 棚橋英樹, 王彩華, 丹羽義典, "移動体ビジョンを指向した小型全方向ステレオシステム(miniSOS)の開発", 第9回画像センシングシンポジウム(SSII03)講演論文集, pp.311-316, 2003.
- [14] 佐藤雄隆, 山本和彦, 桑島茂純, 棚橋英樹, 王彩華, 丹羽義典, "ASTRO-Sensorシリーズの開発とサーベイランスシステムへの応用に関する検討", ビジョン技術の実利用ワークショップ(VIEW2003)講演論文集, pp.166-171, 2003.
- [15] 桑島茂純, 松山方大, "全方向ステレオシステムSOSの設計とその構造", ビジョン技術の実利用ワークショップ(VIEW2005)講演論文集, pp.206-209, 2005.
- [16] 棚橋英樹, 佐藤雄隆, 渡辺博己, 王彩華, 丹羽義典, 山本和彦, "全方向ステレオシステム(SOS)を用いた環境変化の抽出", 第14回外観検査の自動化ワークショップ(VIEW2002)講演論文集, pp.172-177, 2002.
- [17] 佐藤雄隆, 金子俊一, 丹羽義典, 山本和彦, "Radial Reach Filter (RRF)によるロバストな物体検出", 信学論(D-II), Vol.J86-D-II, No.5, pp.616-624, 2003.
- [18] Point Gray Research Inc., "User's Guide and Command Reference - TRICLOPS Stereo Vision Software Development Kit (SDK)."