

特集 ■ユビキタスと VR

ユビキタスステレオビジョン - 実時間実環境ヒューマンセンシング -



依田育士 産業技術総合研究所
YODA IKUSHI



坂上勝彦 産業技術総合研究所
SAKAUE KATSUHIKO

1. はじめに

多地点配置の高速ネットワークに接続したステレオビジョンの利用を前提とした、人間を中心とする(人の存在、その顔やジェスチャ、ハンドサインの)リアルタイム認識技術や、実際に利用する場に適應できるような学習型の認識技術の研究を2000年頃から行っている。ユビキタスステレオビジョンとは、複数のステレオカメラとそれらを利用するアプリケーションという枠組みで構成され、実時間実環境でのヒューマンセンシングを提供するものである。

研究目的は、激しい影の変化などを含む屋内外のどこにでも対応可能なステレオカメラを設置し、そのときに得られる距離情報と画像情報を同時に利用し、いかに役立てるかにフォーカスしている。そのために独立した複数ステレオカメラによる複数対象の認識を前提とし、距離情報から対象領域を着目する手法、対象の記述・認識手法を開発する。我々の研究の中心課題を図1に示した。これに関連する画像認識の研究は既に30年にわたる歴史が存在し、あらゆる手法の開発・適應が試みられてきた。しかしこれほど研究開発が進み、適應範囲が広がっているにもかかわらず、一切制約のない実世界を対象とした画像理解、認識といった範疇に踏み込むと、問題の難易度は飛躍的に高まり、共通のフレームワークすら未だに見えてこないのが現状である。そこで、距離情報を用いることで最も困難な問題である認識対象に着目するためのセグメンテーション問題を劇的に容易にし、その後の認識問題に注力するためのフレームワークを作る。

そもそも RWC プロジェクト後半に参画し [8], 以前から計算機とユーザのより自然な対話を実現するために、基本的にステレオカメラ1台対1人のユーザのための学

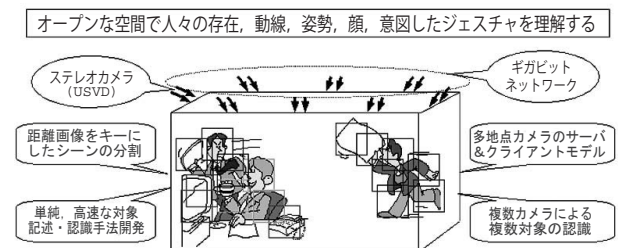


図1 ユビキタスステレオビジョンの概念図

習型ステレオビジョン [7] を研究していた。プロジェクト最終年度を迎えたとき、研究の総括と今後の展開を考慮し、今までの仕組みを複数ステレオカメラと複数人という枠組みに拡張し、総合的なヒューマンセンシングとして捉え直そうと考えた。

ここで我々の研究と「ユビキタス」との位置付けについて述べる。ユビキタスという語は多種多様に用いられているが、インターネットを中心としてユビキタス・コンピューティング [9-11] を包括する意味合いでのユビキタス・ネットワークを主張するものの研究課題である [12]

- (a) 多様なアプライアンスの研究
- (b) ネットワークミドルウェアの研究
- (c) ユビキタス環境におけるサービスの研究

に基づき我々の研究を位置付ける。すると、我々のビジョン研究はその中のアプライアンス、ならびにサービス研究に関連する。これはビジョンセンサがセンサネットワークとして構築されたり、その機能として位置情報を提供したりすることが可能だからである。

位置情報サービスとしては既に GPS や PHS によるものが提供されているが、室内位置など、より高精度なものを求める場合は、IC などのタグを与えるものも考

えられる。しかし、ビジョン技術を用いれば何も持たせることもなくユーザの位置情報を提供できる。さらにビジョンセンサがネットワーク上にユビキタスに配置されることで、位置情報だけでなく、可能な限り個人の詳細な情報を獲得し、物理環境と情報環境を各個人に適合するように自動調節することも可能になる。同様に、安全・安心のためなどに利用することも可能であろう。

以下、我々の提案するユビキタスステレオビジョンとそのアプリケーション事例について述べる。

2. ユビキタスステレオビジョン

2.1 ユビキタスステレオビジョンの基本構成

我々の提案するユビキタスステレオビジョン(USV)[1]とは、表 1 のような特徴を備えているものとした。実時間、実環境、実利用を志向し、あらゆる環境に適応可能、同時に実装を容易にすることを目指している。

そこでまず必要となる 3 次元情報の獲得については、各ステレオカメラに 1 台の PC がローカルに 3 次元形状の復元計算を行い、取得するカラー画像と併せ可能な限りの画像特徴の抽出を行う。我々はこのステレオカメラと PC の組を「ステレオサーバ」と呼んでいる。

次にグローバルな処理として、アプリケーションに合わせて必要なデータや処理特徴を各カメラサーバから別の PC 1 台に統合して全視野内の情報を獲得する。このように必要に応じ、ステレオカメラサーバからデータを収集、統合、利用する側を「アプリケーションクライアント」と呼ぶ。統合には全てのデータを集積する場合や、個々のステレオカメラサーバが認識した結果のみを利用する場合もある。また、両者の中間形態として、必要時にのみデータのみ転送する利用方法も考えられる。

2.2 USV のサーバ&クライアントモデル

ステレオカメラサーバの台数に依存せず、個々のステレオカメラサーバも、全体を統合するアプリケーション

クライアントも稼働するものでなければならない。実際、3 章の室内複数人物認識や、4 章の安全監視事例(踏切空間や駅ホーム端監視)では 4 ~ 12 台で実験を行った。室内と踏切空間内では完全に情報を同期統合することを前提した。複数の情報を統合することで、1 台のカメラが遮蔽、光の差し込みなどで信頼性が低下しても、ロバーストに動作継続することを意図したものである。一方、駅ホーム端などの場合は、基本的に各カメラは弱い同期で動作し、近傍同士のカメラを除き、実時間で統合される情報は画像情報や 3 次元情報から分析された結果のみの非常に少ない記号情報であることを前提としている。

表 2 に、多数カメラの統合と協調方法を基にした分類を示す。以降 3, 4 章で適応される場面は、表のいずれかに分類され、いずれの場合も計算機による自動制御を前提としている。しかし、特に 4 章の安全監視事例において自動判定が難しい場合は、単純にその画像を管理者に自動転送する機能を併せ持つことを想定している。

2.3 ステレオカメラの設置形態と距離情報の処理方法

ステレオカメラの設置方法と、その距離情報の処理方法は、当然のことながらアプリケーションの目的に依存する。しかし、アプリケーションに依存しない共通の距離情報処理方法として、引き出し射影法を提案している[2]。これは、カメラから得られる距離情報をグローバル座標に変換後、空間中に含まれる距離情報を垂直方向に複数段階に分けて射影し、高さ別の空間情報を持った画像群として利用する手法である(図 3 参照)。

このときカメラの設置形態に応じ、階層的に距離情報を処理する手法と、加算的に処理する方法を使い分ける。表 3 にカメラの設置形態とその対象、手法についてまとめた。斜め見下ろし交差型は少数の個人の個別特徴を保持し、認識に用いている。一方、垂直見下ろし型は個人特徴を排除し、大勢の人々に対して集団を分節している。ただし、目的に応じてこれらの手法は混在もする。

表 1 USV の定義

・入力情報は複数のステレオカメラから得られる 3 次元距離とカラー画像
・実時間(リアルタイム)で利用
・実環境で利用(屋内外で利用可能)
・各ステレオカメラはスタンドアロンで稼働
・複数ステレオカメラは静的配置、キャリブレーションは事前に 1 回
・高速ネットワークの利用を基本とするが、低速でも稼働
・動的制御(パン・チルト・ズームなど)をベースとはしない
・切り出しは距離情報を利用し、対象依存モデルを作らない
・認識手法は対象に依存しないものが基本

表 2 USV の統合と独立

i. 強い同期統合型(高速ネットワークで結合)
距離情報、画像情報ともに統合され、信頼性を向上させる。 ex.) 室内空間制御、踏切空間監視
ii. 弱い同期型(低速ネットワークで結合)
距離情報、画像情報はローカルに処理され、必要情報のみ統合される。 ex.) ホーム端安全管理、大空間での動線獲得
iii. 単独型
狭いエリアを 1 台で稼働するスポットインタフェース。 ex.) 重度障害者支援

表 3 ステレオカメラの設置形態と処理手法

A. 斜め見下ろし交差型
対象：テクスチャ情報の利用も含む少人数の精細な認識や高信頼性な認識
手法：段違い引き出し法（情報を重複して加算することで形状を保存）
視野：基本的にカメラ間の視野が重複することを想定
B. 垂直見下ろし型
対象：不特定多数内の個々のトラッキング，人数計測
手法：加算引き出し法（全体形状を安定的に抽出）
視野：カメラ間の視野が一部重複，または独立

表 4 屋内外での USV 利用シーン

a. 室内での人の認識とインタラクション（個人識別も可能）	
・ 人行動ログの取得	・ インタフェースとしての利用
・ 特定作業への支援	・ アミューズメントへの応用
b. 屋外での人の認識（トラッキングが中心）	
・ 安全監視	・ 大空間での動線獲得

2.4 現実の利用シーン

具体的な利用シーンとして，表 4 のような照明の変化が少ない屋内と，激しく光量に変化する屋外での人の認識を想定している。ただし，示した事例以外にも，適応シーンを替えて同様な枠組みでより広いシーンに適応可能と考えている。以下，3 章においては室内の少人数の精細な認識を中心とする複数人物認識とインタフェースについて，4 章においては広域な空間での不特定多数の認識を対象とする利用シーンを中心に事例を説明する。

3. 複数人物認識とインタフェース

ここでは主に室内を中心とした高精度な複数の人々の認識利用例について述べる。既に個人の顔の位置，向き，話者などを識別するシステムは多数存在し，さらに，人間の 3 次元的位置や動きの特定を目的に対象領域を拡げることで全身を撮影するシステムも多数存在する。しかしながらこれらは個々の目的に特化してしまい，顔認識とジェスチャ認識，または動線情報とジェスチャ認識というように本来同時に認識が望まれるような情報を共存させる視点が欠けている。同時に，ユーザは固定的な空間や向きに縛られて認識されている。

そこで，カメラの位置や向きを意識することなく複数の目的を共存可能にするようなフレームワークを提供することを試みる。具体的には室内空間における複数人の個人識別とその全行動ログの取得，および「腕さし」に

よる個人に対応した室内空間の機器操作を可能にするジェスチャインタフェースの実現を目的とした。ここでの腕さしとは，対象物を指示するために，片腕を大きく突き出すことを言う。以下，中心となる人物姿勢認識，個人識別，インタフェースについて紹介する。

3.1 複数人物の個別姿勢認識

実験室において一定領域 (4.5m×3.6m=10 畳) を取り囲むようにステレオカメラ 4 台を四隅に配置した (斜め見下ろし交差型)。そして，その領域内において，

- ・誰がいつシーンに入り，いつ出ていったのか，
- ・その人がシーンの中で何をしていたのか，

ということを実時間で認識することを課題とした。

目標は 10 畳内にいる常識的な複数人 (4, 5 人程度) の立位，座位，伏位の 3 姿勢とその動きの理解である。

図 2 に人の姿勢認識，個人識別，腕さし動作の認識を行っている画面を示した。ここでは，上述の実験室内に立位と座位の人が各 1 名いる状態で，姿勢認識と腕さし認識，さらに顔領域の抽出を行っている。

また，図 3 に階層型引き出し法による射影結果と，それに基づく個人領域の特定，個人の識別結果を示した。

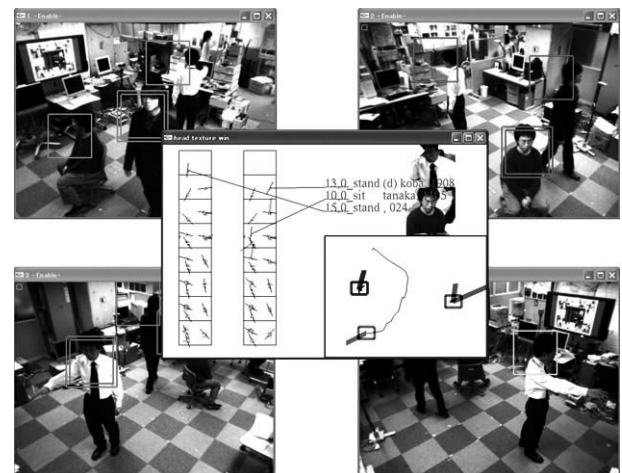


図 2 室内の複数人物認識

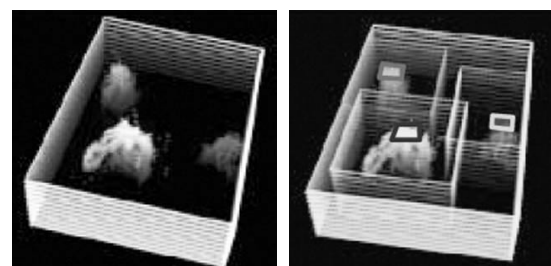


図 3 段違い引き出し法による距離情報利用

左側が階層型引き出し法による射影結果、右側がその結果から各個人の領域を分割、認識した結果である。この例では、立位で腕さしをしている人、立位の人、座位の人の計 3 名がいる。ここでは、個々の顔認識と同時に姿勢と腕さしの有無を識別している。

3.2 顔認識による個人識別

人の状態把握後は個人識別を行う。識別は顔認識を中心とするが、意識してカメラに正面顔を向けた状況で認識を行うのではなく、全く制約なく自然に室内を動く人の顔認識を行う。このために学習用、認識用の顔画像を自然に室内を動く人から取得し、教師なしで正面顔を自動分類して辞書を自動構成し、正面顔らしいと判断した時のみ認識、それ以外の不確かな場合は無視する手法を用いている [3]。

また、実際に人間が行う個人識別は、顔を見ているのではなく、その人の身長や歩き方、普段の服装から判断していると思われる。そこで、常時得られる距離情報と画像情報を利用して、今までのような静的な顔の認識や識別といった枠組みではない個人識別方法の開発も目指している。

3.3 インタフェース

空間内の人の状況、個人識別後は、非接触・非拘束・位置フリーのインタラクションを目標とする。すなわちヒューマンセンシング技術を核にした、コビキタス環境におけるインタフェース機能の開発である。

具体的な目標として、腕さしジェスチャによる室内機器操作インタフェースを実装した。過去の研究事例におけるジェスチャインタフェースは、カメラの向き、場所に依存したものがほとんどである。そこで、室内空間において、位置に依存しないジェスチャインタフェースの実現を目指した。具体的にはユーザの意図した腕さしを距離画像から実時間で獲得する。立位、座位に関係なく、かつ複数ユーザに対応した意図した腕さしジェスチャによる機器操作が可能になっている [4]。また、顔認識と組み合わせることで、個人別に操作の許可や優先権を設定できる。

さらにインタフェースの応用例として、重度障害者のための支援技術にも取り組んでいる。これは電動車いすのコントローラを手で操作できない人に対して、ステレオカメラを車いすに搭載し、顔向き推定 [5] から得られる頭部のジェスチャにより、障害者自身の意思で車いすを操作可能にするインタフェースを提供する試みである。屋内外のどこでも利用可能にすること、実際の臨床を行

いつつ進めるという 2 点に留意しながら取り組んでいる。

この問題は、意図した動作とそうでない動作を見分けるという点において、上述の腕さしジェスチャインタフェースとは極めて類似点が多い。我々の研究は環境に一切の制約を加えない。同時に、操作者には意図したジェスチャをしてもらうが、それ以外の普通のジェスチャに制限を加えないで行っている。決められたジェスチャのみ行うことを前提とした認識システムや、ジェスチャ認識ゲームと、我々のスタンスは根本的に異なる。

4. 屋外大規模空間での状態監視 (安全管理への適用)

上述の室内のより詳細なヒューマンセンシングとは異なり、個々の人のジェスチャ、顔認識などを排除した不特定多数の状態監視について述べる。特に屋外など設置場所を一切選ばない広い空間に適応させる。この目的の応用事例として、踏切内監視、駅ホーム端監視、イベント空間監視などについて継続的に実験を行っている。

これらの実験は実環境において、どの程度の解像度画像によって何が可能なのか。必要となるアルゴリズムは何なのか。実環境に耐え得るハードウェアスペックはどの程度なのか。これらを実地試験で獲得するために行っている。表 5 に、今までに行ってきた主要実験例を一覧した。以下個々の目的別に実環境実験について紹介する。

4.1 踏切空間監視

現在の踏切内安全監視装置には、複数の赤外線センサが利用されている。この方式では赤外線センサのライン上に障害物があった場合のみ反応するので、自動車は確実に検知できても、人や自転車などがそのラインの間にいた場合は、全く無力なセンサとなっている。そこで、ステレオカメラからの距離情報を利用して、踏切内を認識する研究も始まっている [13]。この事例では、単体のステレ

表 5 継続的な広域空間監視実験概要

	日時	場所	目的	カメラ台数
踏切	2003. 1-2	東急東横線 元住吉駅踏切	踏切空間の人の監視	6 台 斜め見下ろし交差型
	2004. 6	東急東横線 自由が丘駅踏切	踏切空間の人の監視	4 台 斜め見下ろし交差型
駅 ホ ー ム	2001. 11	JR 牛久駅 3 番線ホーム	ホーム端端監視実験 転落状況のデータ取得	4 台 垂直見下ろし型
	2001. 11	東急東横線 横浜駅 1 番線ホーム	ホーム端監視実験 乗客のデータ取得	5 台 垂直見下ろし型
	2003. 1-2	東急東横線 元住吉駅 3 線ホーム	ホーム上の人の監視 車いすの認識 白杖の認識	1 台 斜め見下ろし型
広域	2005. 3-9	愛知万博 グローバルハウス オレンジホール	会場内部の人数計測 行動軌跡の解析 RF-ID との連携実験	7 台 垂直見下ろし型



図4 USV による踏切空間監視

カメラにより精度を安定させることは行われているが、複数台による精度向上、安定性確保は行われていない。

USV による監視方法では、踏切の四隅に踏切空間を取り囲むようにステレオカメラを設置し（斜め見下ろし交差型）、4 台のステレオカメラで完全に空間を監視する。図4の四隅の画像が、4 台のステレオカメラから得られたカラー画像であり、中央は統合された3次元情報を示している。このような構成で、単体の必要解像度、太陽光の影響、設置角度の影響を、また複数台化の効果などあらゆることを評価している。

4.2 駅ホーム端安全管理

駅のホーム端からの転落事故による死亡者は年間数十件に達し、特に都市部での鉄道、地下鉄においては緊急に解決すべき課題となっている。既存のセンシング技術としては落下検知マットが利用され、超音波、赤外線センサなども検討されているが、検知範囲や誤検出への対応に難がある。そこで、ステレオカメラを駅ホーム端の天井から下を見るように直線に並べ（垂直見下ろし型）、ホーム端の人々の状況を監視する[6]。実際に東急東横線横浜駅でカメラから取得した3次元画像を図5に示した。

現実の複雑な状態の人々を対象に、ホームからのデータ転落実験をJR牛久駅で、ホーム端の軌跡獲得実験を東急東横線横浜駅で、ホーム上の人々の自動分類に着目した実験を元住吉駅で行い、データを取得した。

JR 牛久駅ではホーム上と線路面上両方で、人の落下、座り込み、寝込み姿の3次元情報から、全時間帯における（夕日の差し込み時、夜間を含む）人の存在を確実に



図5 USV によるホーム端3次元復元図

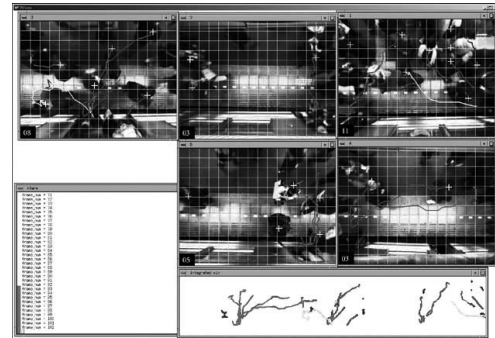


図6 USV による全体の動線管理

検知できることを確認した。

横浜駅においては、始発から終電までの全時間帯において、実際の乗客の出入りを撮影した。図6には、全乗客の動線管理を示した。各カメラ単位に処理された僅かな結果情報から全体の動線を作成する弱い同期統合型のシステムとなっている。この処理戦略により、今後カメラが数十台単位で増加しても対応可能になっている。

4.3 広域空間監視

上述の監視技術は、踏切やホーム端といった安全監視だけの目的以外には、より大きな空間での動線計測、監視に適用可能である。そこで現在開催されている愛知万博（愛・地球博）のグローバルハウス・オレンジホール（図7）において半年間の常設実験（垂直見下ろし型）を行っている。実験の目的は以下の三つである。

1. リアルタイムの大きな領域単位での人数計測
2. 視野内の人の行動分析
3. RF-ID タグ (Aimulet) との連携実験

あらゆるセンシングには向き、不向きがあるので、他のセンサとビジュアルセンサをどのように組み合わせるのかなどへの研究の発展、展開は無限の可能性があると考えている。それに向けてのグランドチャレンジが開始されたところである。

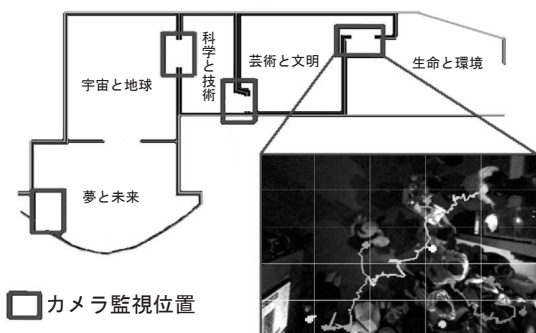


図7 USV 愛知万博グローバルハウス・オレンジホール

5. まとめと今後の研究

USV の基本構成, ならびに現実のアプリケーション事例について述べた. 本研究の目的は, 対象に応じ, ステレオカメラを静的に適宜配置することにより, そのシーン内の必要な情報を掌握するために, 3 次元情報とカラー画像情報を同時に利用する手法の研究開発である.

具体的なアプリケーションとしては精細な認識を求める主に室内での生活支援技術と, 大規模空間を対象とする主に屋外での安全性向上支援の事例を示した. それぞれの目的に合わせて複数のステレオカメラが, 強い同期で動作, 弱い同期で動作, あるいは単純にシングルで動作することができる.

現在, 最重要事項として, 現在までの全ての実験結果を反映させたハードウェアであるユビキタスステレオビジョンデバイス (USVD) を完成させた. これはステレオ距離計算部分を全て受け持つハードウェアを実装したハーフサイズ PCI ボードと, 専用ステレオカメラから構成される. これにより, 計算機本体の CPU を全て認識処理に使うことを可能にする. さらにランチボックスサイズのセットトップボックス化をも可能にし, どんなどころにでも設置できるハードウェアになる. 今後は, この USVD をベースとして実装の展開を図る. 同時に, その中で必要となる人を中心とした対象の自動的な学習方法, 対象の記述方法, 現実シーン内の認識方法の開発を進める.

参考文献

- [1] I. Yoda and K. Sakaue: "Concept of Ubiquitous Stereo Vision and Applications for Human Sensing" Proceedings IEEE Inter. Symposium on Computational Intelligence in Robotics and Automation, pp.1251-1257 (2003)
- [2] I. Yoda, et al.: "Human Body Sensing Using Multi-Point Stereo Cameras" Proceedings of Inter. Conf. on Pattern Recognition 2004, Vol.4, pp.1010-1015 (2004)
- [3] I. Yoda, et al.: "Automatic Face Classification and Recognition Using Self-Organization" Proceedings of Inter. Conf. on Pattern Recognition 2004, Vol.4, pp.1006-1009 (2004)
- [4] Y. Yamamoto, et al.: "Arm-Pointing Gesture Interface Using Surrounded Stereo Cameras System" Proceedings of Inter. Conf. on Pattern Recognition 2004, Vol.4, pp.965-970 (2004)
- [5] B. Raytchev, et al.: "Head Pose Estimation by Nonlinear Manifold Learning" Proceedings of Inter. Conf. on Pattern Recognition 2004, Vol.4, pp.462-466 (2004)
- [6] 依田, 細谷, 坂上: "ユビキタスステレオビジョンによる駅ホーム端安全管理" 電学論 C Vol.124, No.3, pp.805-811 (2004)
- [7] I. Yoda and K. Sakaue: "Utilization of Stereo Disparity and Optical Flow Information for the Computer analysis of Human Interaction" Machine Vision and Applications, Vol.13, No.4, pp.185-193 (2003).
- [8] 産総研実世界知能研究班: "産業技術総合研究所における RWC 研究発表資料集" 2000 (2001), 2001 ~ 2002 (2002)
- [9] M. Weiser: "Some computer science issues in Ubiquitous Computing" Communications of the ACM Vol.36, No.7 pp.74-84 (1993-07)
- [10] TRON PROJECT Official HP: <http://www.tron.org/>
- [11] J. Wejchert: "「姿を消すコンピュータ」: 進行中の研究プロジェクト" 情処誌 Vol.43, No.8, pp.880-883 (2002-08)
- [12] 森川, 南, 青山: "ユビキタスネットワークへの道" 情処誌 Vol.43, No.6, pp.631-638 (2002-06)
- [13] 柴山, 埴, 太田: "踏切障害物検知用のステレオ画像装置の開発" 第 9 回画像センシングシンポジウム 予稿集, pp.1-6 (2003)

【略歴】

依田育士 (YODA Ikushi)

(独) 産業技術総合研究所

情報技術研究部門 主任研究員

1992 年東京都立科学技術大学大学院工学研究科修士課程修了. 同年電子技術総合研究所入所. 入所後, 画像処理手順の自動学習法や, 人を中心とした実時間認識システムに関する研究など, コンピュータビジョンや画像理解の研究に従事. 筑波大学連携大学院助教授 (併任). 工学博士.

坂上勝彦 (SAKAUE Katsuhiko)

(独) 産業技術総合研究所

情報技術研究部門 部門長

1981 年東京大学大学院工学研究科博士課程修了. 同年電子技術総合研究所入所. 画像処理、コンピュータビジョンなどの研究に従事. 筑波大学連携大学院教授 (併任). 工学博士. 1979 年度電子通信学会学術奨励賞受賞. 1985 年度情報処理学会論文賞受賞.