

分散リアルタイムセンシングによる 高速動作獲得技術

東京大学大学院工学系研究科計数工学専攻

鏡 慎吾 swk@k2.t.u-tokyo.ac.jp

東京大学大学院情報理工学系研究科システム情報学専攻

石川 正俊 ishikawa@k2.t.u-tokyo.ac.jp

コンピュータ技術・ネットワーク技術を実世界での人間支援に活用しようとする研究が盛んに行われている。適切な支援サービスを提供するためには、実世界の情報を適切に獲得できることが必要であり、そのためには、センシング対象・センシング目的に応じた適切なセンシングを実現することが重要となる。本稿ではこのような視点から、センシング特性のうち特に視覚情報のリアルタイム性に着目し、筆者らの研究室で行っているディジタルビジョンチップとその応用に関する研究を中心に概説する。またそのリアルタイムネットワークへの展開について述べ、その課題について議論する。

◆人間支援のためのセンシング技術

飛躍的な進歩を続けているコンピュータ技術・ネットワーク技術を、実世界における人間の活動の支援に利用しようとする動きが盛んになっている。Smart Rooms¹⁾、Robotic Room²⁾ 等さまざまなコンセプトが提案されており、いずれもセンサ・アクチュエータやプロセッサを環境中に配置、あるいは人間に装着し、それらを利用して人間のおかれた状況に応じた適切なサービスを提供することを目指しているという点で共通している。このような技術の実現を考える際、実世界からの情報の入口にあたるセンシング技術は重要な役割を受け持つ。適切な支援サービスを提供するためには、人間自身の様子や周辺環境の状況、さらにはそれらの相互関係についての情報を適切に効率よく取得する必要がある。

特に、対象の情報を非接触で得ることのできる視覚センシングは、センシング対象を拘束せずに多くの有用な情報を取得できる点で、大きな期待を集めている。中

でもコンピュータビジョンの文脈では、分散協調視覚³⁾やVSAM (Video Surveillance and Monitoring)⁴⁾ などのような大規模なプロジェクト研究も行われ、大きな成果を上げている。その一方で、センサ情報の利用目的が広範になり、要求が高度化するのにつれて、従来の視覚センサの性能では必ずしも満足できないケースが顕在化するようになってきた。

一般にセンシングあるいはセンサ情報処理は、利用者にとって有用な情報を実世界から選択・抽出する行為を意味する⁵⁾。有用な情報とはアプリケーションによって千差万別であり、何をセンシングの対象とするか、その情報に基づいて何を行おうとしているのか等に応じて、抽出すべき情報の種類や質、量が異なってくる。

視覚センシングシステムとしては、従来より CCD 型のイメージセンサと、そのビデオ信号出力を受け取って処理する何らかのプロセッサの組合せが広く用いられている。しかしこれらのシステムでは、得られる視覚情報はある一定のフレームレート以下に、また明るさに関してある一定のダイナミックレンジ以内に制限されてしまっている。人間の活動を支援するためのセンシングを考えた場合、人間の動作獲得やその周囲における突発的な危険検出などのために高速な運動の認識が必要となるケースや、さまざまな照明条件下での運用が必要となるケースなど、多くの場合にこの制限は大きな問題となる。そしてこれらの制限は、プロセッサへの入力の時点ですでに生じてしまっているため、プロセッサにおける処理の如何にかかわらず、必要な情報を完全に復元することはできない。

このような問題を解決するためには、2つのアプローチが考えられる。1つは、あらゆる性能指標（速度、感度、精度、測定レンジ、など）において考え得る最大の

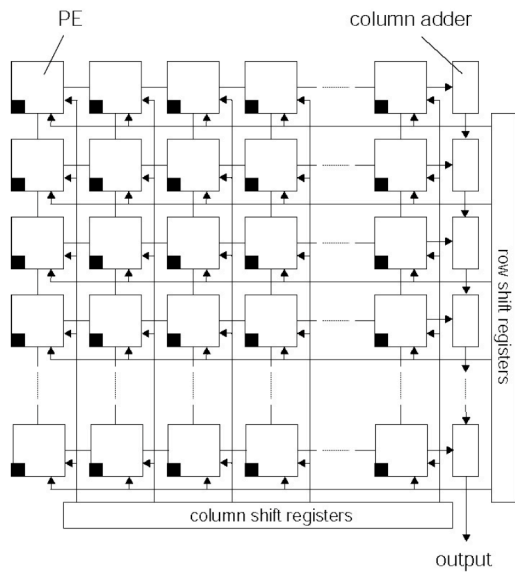


図-1 ビジョンチップの全体構成

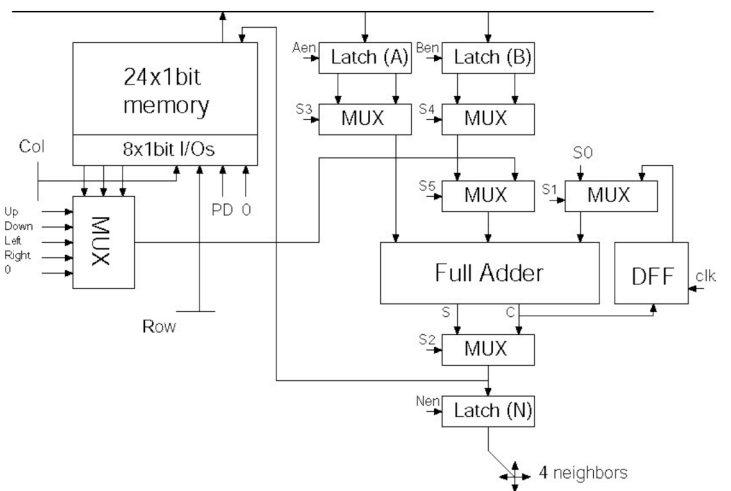


図-2 ビジョンチップのプロセッシングエレメント構成

性能を発揮するセンサを導入するという方向である。しかしこれらの間にはトレードオフの関係があることが多く、また仮にすべてを満たせたとしてもコスト面で非現実的な解となってしまうかねない。これに対するもう1つのアプローチは、これらの各性能をセンシング目的や環境に適応させることのできるセンシング構造を導入するという方向性である。

本稿では、この後者のアプローチによって、センシング対象・センシング目的に応じた真のリアルタイムセンシングを実現するためのアーキテクチャについて述べる。まず、高速かつ柔軟な視覚処理を実現するデジタルビジョンチップとその応用について紹介し、リアルタイムネットワークへの展開とその課題について論じる。

◆高速視覚センシングとビジョンチップ

各画素に光検出器（PD; Photo Detector）とともに何らかの処理回路（PE; Processing Element）を集積し、画素数のレベルの並列度での並列処理を行うイメージセンサをビジョンチップと呼ぶ。ビジョンチップに関する研究の多くはアナログ処理に基づくものであるが、近年ではデジタル回路をPEとして採用するデジタルビジョンチップの研究も進められている。ビジョンチップの利点の1つは、ビデオ信号によるフレームレートの制限にしばられることなく、高速な視覚処理を実現できる

点にある。デジタルビジョンチップはさらに、処理の内容をソフトウェアとして記述できるため、多種多様な要求に応えることができるという利点を持つ。

石川らは、各画素にメモリとALUからなる汎用のPEを集積するデジタルビジョンチップの研究・開発を行っている⁶⁾。ここではその最新の開発例⁷⁾を紹介する。

センサ全体の構成を図-1に示す。全画素が一度に同一のインストラクションを実行するSIMD（Single Instruction stream, Multiple Data stream）型の並列処理を採用している。図-2に示すように各画素にはビットごとにランダムアクセス可能な24ビットのメモリと、ビットシリアルALUが備わっており、1ビットずつの演算を繰り返すことで多ビット演算や入出力を含む汎用の処理を実行する。

画素間の通信は、上下左右の4近傍接続を基本としているが、複数のPEを連結する機能を備えているため、局所的なパターン処理だけでなく、たとえば画像の面積・重心といった大域的な特徴の計算も効率よく実行できる。そのため、2次元パターン情報からスカラ、あるいはベクトル情報までデータ量を圧縮してチップ外に取り出すことが可能となる。

また、視覚処理のアルゴリズムは必ずしも画素数レベルの並列度を持つとは限らないため、画素ごとの完全並列処理構造はかえって非効率になってしまう場合があ

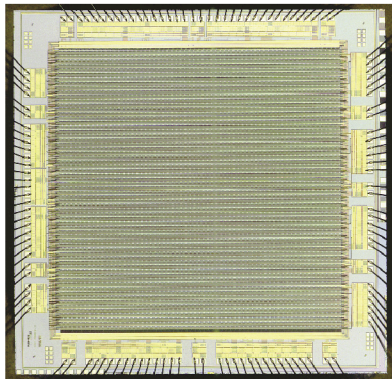


図-3 試作チップの写真

る。このアーキテクチャは、PE 連結機能を用いて複数の PE を任意の粒度でブロック化できるため、処理内容に応じてハードウェアを再構成可能とすることで、この問題に一定の解決を与えている。

このアーキテクチャに基づいて試作されたチップの写真を図-3 に示す。0.35 μ m CMOS プロセスにて、エリアサイズ 5.4mm × 5.4mm 内に 64 × 64 画素が集積されている。PE あたりの面積は 67.4 μ m × 67.4 μ m となっている。

◆ビジョンチップによるリアルタイム視覚処理

リアルタイム視覚処理は一般に、1 フレームごとにデッドラインが訪れるリアルタイム処理であるとみなすことができる。1 フレームの長さをどうとるべきか、またそのデッドラインがハードデッドラインかソフトデッドラインかといった仕様は、アプリケーションに応じて、すなわちセンシング対象の運動の状態や、その視覚情報をどのように利用するか（タスクのプランニングに利用するのか、機械系の制御に利用するのか等）等によって定義されるべきものである。

前章で述べたデジタルビジョンチップではビデオレートの制限を受けずに視覚処理を記述できるため、フレームレートをアプリケーションに応じて自在に引き上げることができる。ただしそのためには、特にフレーム時間内に確実に処理を完了するハードリアルタイムが要求される場合には、SIMD アレイであるビジョンチップへのインストラクション供給がリアルタイム性を保って行われることが必要となる。

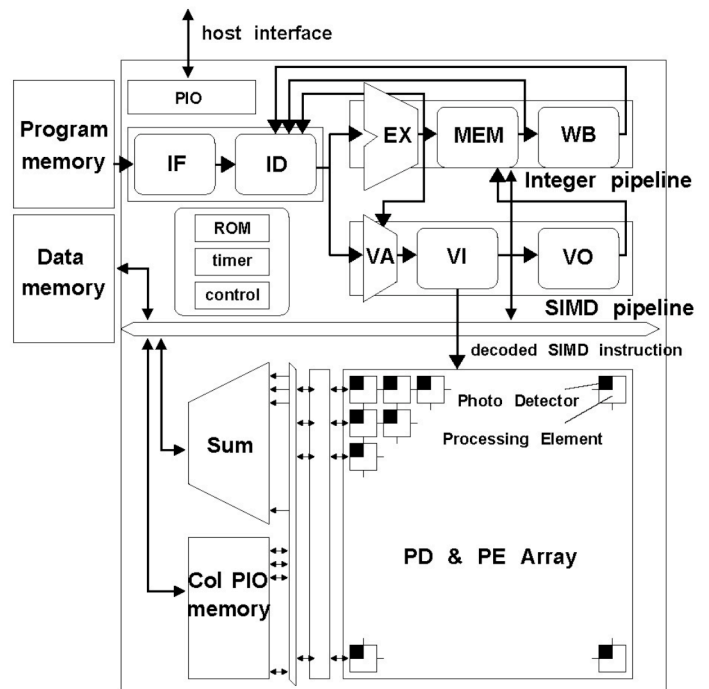


図-4 ビジョンチップの制御アーキテクチャ

また、視覚センシングにおいて要求される性能は、高速性だけではない。すでに述べたように、ダイナミックレンジや感度特性などのセンシング特性も、さまざまな要求に適応させることが必要となる。光検出を行う物理構造である PD とソフトウェアの実行主体である PE が、画素ごとに直結しているデジタルビジョンチップの構造は、この点に関しても柔軟な適応を可能とする。たとえば明るい光でもセンサの出力が飽和しないように、光電効果によって発生する電荷を蓄積する時間をソフトウェアでうまく調整するといったことが可能となる。ただしそのためには、フレーム時間の制約から生じるリアルタイム性よりもさらに厳しいリアルタイム性を満たした SIMD アレイの制御がなされなくてはならない。

これを実現するため、専用の制御アーキテクチャが提案されている。その構成を図-4 に示す。このアーキテクチャは、プログラム制御を外部のホストに依存せずに行えるよう、RISC 型プロセッサとして完全に独立して動作する。SIMD アレイのインストラクションの生成・供給のフローを完全にパイプライン化し、かつ整数パイプラインとの相互作用も含めた全体として動的なストールが発生しない構造をとっている。この構造により、十分な制御レートを確保すると同時に、命令サイクルの粒

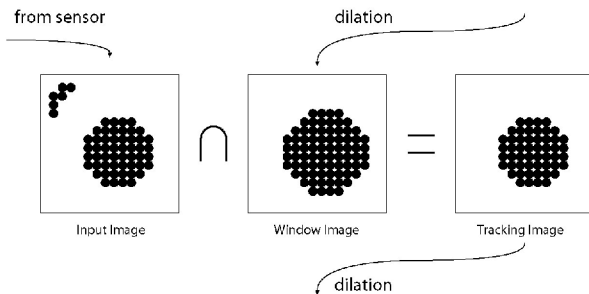


図-5 セルフウィンドウ法のアルゴリズム

度でのリアルタイム性を保証している。

この高いリアルタイム制御能力を利用して入力光の A-D 変換の特性を制御する、ソフトウェア A-D 変換と呼ぶ手法も提案している。この手法は、任意に与えられた特性をノイズを最小化する条件の下で実現するために、A-D 変換の各しきい値について PD の動作タイミングと入力電圧のそれぞれをソフトウェアで適切に制御するものである。

このようにデジタルビジョンチップおよびそのリアルタイム制御系からなるセンシングシステムは、アプリケーションが必要とする情報を、それが失われるより前の段階で効率よく抽出することを可能とする。それと同時に、その過程がソフトウェアで記述されているため、多様な要求に応えることが可能となっている。すなわち、センサと処理部を一体化することによって、アプリケーションからの個々の要求を、物理的なデバイスからソフトウェアとして分離可能な構造を実現するものであるといえる。このようなソフトウェアで制御される適応機構を効果的に実現させるためには、今まで以上に物理的特性を巧みに利用したデバイス設計が必要になる。

◆高速視覚による人間支援

センシング目的に合致した真のリアルタイムセンシングが実現されることによって、視覚センシングの応用範囲は大きく広がる。特に、従来の視覚処理では対応できなかった高速な対象の認識が可能となる。これには、従来は取得できなかった情報が取得できるようになるという単純な理由の他に、高速化によって視覚処理がきわめて簡単化されるという理由が大きな役割を果たしている。



図-6 ジェスチャ認識への適用例

フレームレートが高くリアルタイム性が高いことは、対象の変化を完全に把握できることにつながる。これは、フレーム間で対象に関する画像の差分が微小となることによる。特に 2 値画像においては、任意の有限の速度で変化する対象に対してある有限のフレームレートが存在して、そのフレームレート以上では「フレーム間で画像の全要素がただか 1 ピクセルしか移動しない」ようにすることができる。

この特徴を利用することで、フレーム間の画像マッチングのように従来は計算量の大きな処理が必要とされたタスクを、ごく簡単な処理のみで実現することができる。これを行うセルフウィンドウ法と呼ぶアルゴリズムを図-5 に示す。このアルゴリズムは、追跡対象の画像が、対象自身の 1 フレーム前の画像を 1 画素だけ膨張した範囲の中に常に含まれることを用いたもので、画像膨張演算と論理積演算のみで注目している対象のみを抽出することが可能となっている。

このアルゴリズムを用いて、2 軸のサーボモータ上のカメラで対象を追跡するシステムが実現されている。このシステムでは、上記のアルゴリズムによって目的とする対象だけを抽出し、それを追跡するようにカメラを制御する。視覚を含んだ制御ループが 1kHz のフィードバックレートで動作しており、高速で不規則に動く対象でも確実に追跡することが可能となっている。

このような対象追跡タスクの応用として、ジェスチャ認識が挙げられる。人間の動作は、野球の投球動作を考

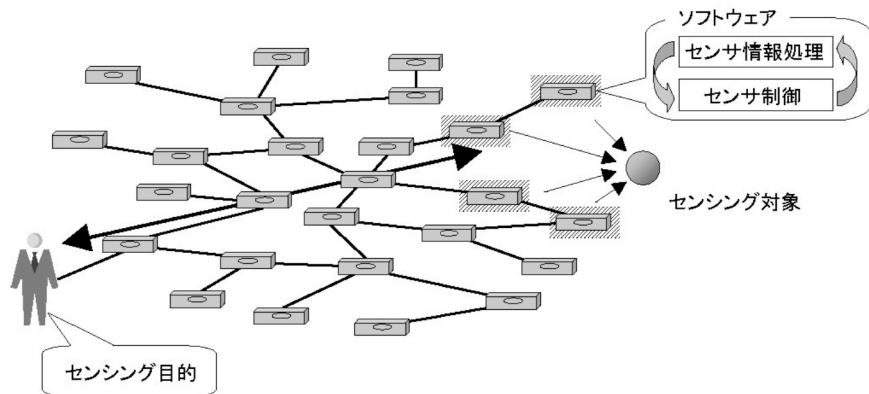


図-7 Reconfigurable Sensor Network の概念

えてもたかだか 150km/h 程度である。フレームレートを 1kHz にとれば、フレーム間で 42mm 程度しか移動しない。そのため、人間の動作の完全な獲得が実現できる。例として、前述の対象追跡システムにて人間の手の動きを獲得している様子を図-6 に示す。このように人間の行動を獲得することや、それをウェアラブル機器や情報家電への入力インタフェースとして利用することなどが可能となると期待できる。また、同様に考えることで、人間の周囲の環境の変化や、突発的に発生する事象の検出などへも応用できると考えられる。

◆センサネットワークへの展開

本特集の他の記事でも解説されているように、生活環境や社会のインフラストラクチャに埋め込まれたリアルタイムネットワークの研究・開発が盛んに行われている。ネットワーク技術によって膨大な数のセンサが互いに接続されたとき、センサとアプリケーション、センサと利用者との関係は、従来の固定的な関係から多対多、かつ動的な関係へと変貌する。

このようなセンサネットワークに関する研究が近年活発化している。デバイスの面からは、たとえば Smart Dust⁸⁾ や WINS⁹⁾ などのプロジェクトで、小型化や低消費電力化に主眼をおいたセンサノードの開発が進められている。またネットワークの面からは、モバイル Ad-hoc ネットワーク技術を中心とした議論が活発に行われていると同時に、よりセンサネットワークに適したかたちを模索する、新たなネットワークアーキテクチャの研

究も行われている¹⁰⁾。

前章までで議論してきたように、センサと処理部を一体化することによって、アプリケーションからの個々の要求を、物理的なデバイスからソフトウェアとして分離可能とする構造が実現される。この構造は、ネットワーク技術との組合せを考えた際に、重要な意味を帯びることになる。

ソフトウェアは、本質的に可搬性を有する。すなわち、あるアプリケーションに対応していったん構成されたソフトウェアは、保存や複製、再利用が容易に行える。ネットワークを通じて他の場所に転送することもできる。この特質を利用すれば、何らかのセンシング目的を持ったときに、そのセンシング対象の情報を効率よく得るために最も適したデバイスをネットワーク中から選び、そのデバイスに、最適なセンシング特性を記述したソフトウェアをロードすることで、所望の位置に、所望のセンサを任意に「生成」するのと等価なことが可能となる。図-7 に示すこの概念を、ここでは Reconfigurable Sensor Network と呼ぶ。

たとえばある特定の人間の動作を取得し続けようとする場合は、人間が移動するのに合わせてデバイスからデバイスへとソフトウェアがマイグレートしていく。それと同時に、センサと人間の相互位置関係や照明条件、運動状態に応じて、各センサにおけるセンシング特性を適応させなければならない。この考え方は、ユビキタスコンピューティングが目指す、環境中に偏在する潜在的な計算能力を自由に利用するという概念を実世界の中のセンサおよびセンサ情報処理に拡張したものとみることが

できる。これの整備により、環境中に偏在する潜在的なセンシング能力をソフトウェアによって活性化して自由に利用できる環境の実現が期待される。

これと同様にセンサノードにプログラマビリティを持たせることを目指すものとして、SensorWare¹¹⁾やSQTL¹²⁾の研究が挙げられる。前者¹¹⁾はセンサ情報処理を汎用スクリプト言語 Tcl で記述してこれをセンサノードで実行するものであり、スクリプトには近隣ノードに自己を複製する能力を持たせている。後者¹²⁾では、SQTL と呼ばれる専用言語を導入し、センサノード間で協調してデータの問合せや処理の指示を行うことを可能としている。Reconfigurable Sensor Network もこれらと同様の考え方に基づいているが、センシングを実現する物理構造にまで踏み込んだ適応を視点に入れている点が異なっている。

一方でこのような環境を現実のものとするためには、大きな課題が山積している。Estrin ら¹⁰⁾はセンサネットワークにおける課題として、膨大な数のセンサノードの無人運転が必要な点と、環境やタスクの動特性への適応が必要な点を挙げている。

目的に合致したセンシングを実現するためには、多数のセンサ群の中から必要な情報を得ることのできるセンサ集合を特定し、ノードでの消費エネルギーや通信帯域、通信遅延などのさまざまなコストを考慮しながら、それらからの情報を統合することが必要となる。たとえばアプリケーションが高いリアルタイム性を要求している場合、いかに精度のよいセンシングのできるセンサノードが存在しても、そのノードからの情報伝達に大きな遅延があったとすると、すべて台無しになってしまう。これらの各コスト間のトレードオフを定量的に扱った最適化を実現することが重要であり、また同時に、演算処理やアクチュエーションのなされる位置の最適化も含んだ包括的な問題として考察することが必要となると考えられる。

また、センサノードで行われる処理を記述するためのソフトウェアアーキテクチャも重要な課題となる。Reconfigurable Sensor Network ではセンシングの物理構造に踏み込んだソフトウェア制御を考えているため、低レベルなデバイス制御を何らかのかたちでソフトウェアに開放することが要求される。この点が、軽量のスク

リプト言語の実行のみを想定している既存研究^{11), 12)}との大きな違いであり、そのためのオーバーヘッドや安全性、相互接続性を考慮した慎重な検討が必要となる。

◆まとめ

本稿では、人間支援のために重要となるリアルタイムセンシング技術について、特に筆者らの研究室で行ってきたデジタルビジョンチップとその応用に関する研究を中心に概説し、またそのリアルタイムネットワークへの展開について述べた。実環境の情報を適切に取得するためには、センシング対象・センシング目的に応じて、真に必要なスペックでのセンシングを実現することが必要であり、そのためには既存のセンサデバイスに関する固定観念をふりはらって、本当に必要なものは何かを考えることが重要である。今後もこのような観点から、真に必要なとされるセンシングアーキテクチャを追求していきたい。

参考文献

- 1) Pentland, A.: Perceptual Intelligence, Communications of the ACM, Vol.43, No.3, pp.35-44 (2000).
- 2) 森 武俊, 佐藤知正: 人間支援システム—ロボティックルーム, システム/制御/情報, Vol.44, No.3, pp.151-156 (2000).
- 3) 松山隆司: 分散協調視覚プロジェクト, 日本ロボット学会誌, Vol.19, No.4, pp.416-419 (2001).
- 4) Collins, R. T., Lipton, A. J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., Burt, P. and Wixson, L.: A System for Video Surveillance and Monitoring, CMU-RI-TR-00-12, Technical Report, Carnegie Mellon University (2000).
- 5) 山崎弘郎, 石川正俊編著: センサフュージョン, コロナ社, 東京 (1992).
- 6) 石川正俊, 小室 孝: デジタルビジョンチップとその応用, 信学論(C), Vol.J84-C, No.6, pp.451-461 (2001).
- 7) 石川正俊, 小室 孝, 鏡 慎吾: デジタルビジョンチップの新展開, 信学技報, ICD2002-39 (2002).
- 8) Kahn, J. M., Katz, R. H. and Pister, K. S. J.: Next Century Challenges: Mobile Networking for "Smart Dust", Proc. Ann. ACM/IEEE Int. Conf. Mobile Computing and Networking, pp.271-278 (1999).
- 9) Asada, G., Dong, M., Lin, T. S., Newberg, F., Pottier, G., Kaiser, W. J. and Marcy, H. O.: Wireless Integrated Network Sensors: Low Power Systems on a Chip, Proc. European Solid State Circuits Conf. (1998).
- 10) Estrin, D., Govindan, R., Heidemann, J. and Kumar, S.: Next Century Challenges: Scalable Coordination in Sensor Networks, Proc. Ann. ACM/IEEE Int. Conf. Mobile Computing and Networking, pp.263-270 (1999).
- 11) Boulis, A. and Srivastava, M. B.: A Framework for Efficient and Programmable Sensor Networks, Proc. IEEE Int. Conf. Open Architectures and Network Programming (2002).
- 12) Jaikaeo, C., Srisathapornphat, C. and Shen, C.: Querying and Tasking in Sensor Networks, SPIE's Ann. Int. Symp. on Aerospace/Defense Sensing, Simulation and Control (2000).

(平成 14 年 12 月 11 日受付)