

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 4 月 30 日現在

機関番号：22604

研究種目：基盤研究(C)

研究期間：2010 年度 ～ 2012 年度

課題番号：22500162

研究課題名（和文）確率的処理を活用した屋外での低演算量高精度シルエット抽出とその追跡

研究課題名（英文）Low operational amount silhouette extraction and its application to video understanding based on statistical processing

研究代表者

西谷 隆夫（NISHITANI TAKAO）

首都大学東京・システムデザイン研究科・教授

研究者番号：00389206

研究成果の概要（和文）：

本プロジェクトは電池駆動のロボットやモバイル端末等に視覚を与えるためのもので、従来のものと比べると演算量で1ケタ以上少なく、かつ、安定性を増すアルゴリズムとして開発できた。視覚を与えるにはカメラからの映像から安定した物体の形（以降シルエット）を抽出する必要がある。このためには動くものを抽出する前景分離に対して画像強調と影除去法が必要である。画像強調では従来の目の反応を画像の一般的性質を用いた近似解とすることで演算量削減を実現した。また、影除去では従来の概念と異なる空間スペクトルを用いた除去法を提案し目標を下回る演算量で実現した。この結果として得られたシルエットで歩容解析も行い、好ましい結果が得られた。また、前景分離 LSI を FPGA により 150mW、ASIC 設計で 30mW の消費電力で実現できることを示した。

研究成果の概要（英文）：

The research project is to establish an only 10% operation amount sight/vision function, compared to the conventional one, in order to give the sight capability to robots as well as mobile terminal, because these applications should be powered by battery cells. The sight/vision fundamental function is based on the extraction of silhouette of person in a scene. This fundamental function is composed of the clear separation of foreground objects, followed by the elimination of shadows from the foreground objects; both of the functions require tremendous operations in the conventional approach. In order to achieve the target performance, new modification of Multi-scale Reinex and a novel shadow assumption have been employed. The resulted silhouette has also been applied to walk identification trial and shows the effectiveness. In order to show the possibility of battery support applications, foreground segmentation processor has been developed: the result of power dissipation is 150mW for FPGA, equivalently 30mW for ASIC design.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010 年度	1,300,000	390,000	1,690,000
2011 年度	1,000,000	300,000	1,300,000
2012 年度	800,000	240,000	1,040,000
年度			
年度			
総計	3,100,000	930,000	4,030,000

研究分野：総合領域

科研費の分科・細目：情報学・知覚情報処理、知能ロボット

キーワード：画像強調、前景分離、影除去、低演算量、アーキテクチャ

1. 研究開始当初の背景

ロボットなどにカメラを用いて視覚を与えて人とコミュニケーションを行うためには、人を抽出してその人の動作認識が行なえる必要がある。このために用いられる主要な技術は前景分離と言うもので、映像の中から動いているものを抽出する方法である。しかしながら従来の方法は厳しい条件では測定されていない。

前景分離は入力フレーム画像から背景画像を引き去る背景差分法を用いるが、この場合2つの問題点がある。第1はカメラ画素のRGB成分が8bitと狭いため、人間には背景と前景の区別できてもカメラを通すと区別できない場合が多い。また、第2の問題点は影の部分も背景の明るさが変化するため前景となることである。また、従来法は全て画素ごとの処理であり、最近の携帯電話で用いられているカメラは200万画素程度のHDTVに近づいており、また、人の動きを認識するためにはビデオのフレームレートに合わせてシルエットを抽出したい。これ等の状況から、Tera FLOPS並みの演算量が必要となり、現状では数百ワットの消費電力を必要とするGPUコンピューティングでやっと到達できる。つまり、電池駆動のロボットの視覚を実現するには、第3の問題点として演算量を削減し、低消費電力での実現方法が明示されなくてはならない。

2. 研究の目的

屋内外で活動できるロボットの視覚としてシルエット抽出法を確立する。このため、確実な前景分離法と影除去法を従来法の10%以下の演算量で実現できるアルゴリズムを導出し、これを入力映像からコンテンツ理解までの総合特性で評価する。

確実な前景分離を行うためには前景と背景の画素値の差が小さくなった場合、区別が付きにくい。前景分離に向けた画像強調法が不可欠である。また、前景分離では背景は微動するが、大きな変化はないと仮定している。このため、動くものを分離することである。この場合、日中では動きとともに影が現れる。この影はコンテンツを理解するうえで不要なものであり除去する必要がある。前景分離領域から影部分を除去したものがシルエットになる。今回は安定した影除去を行ってコンテンツ理解までを見据えた個人認識の精度でも判断したい。

さらに、ロボットの視覚や携帯電話に活用するためには電池駆動になる。低消費電力による実現法が不可欠である。これまでの研究で前景分離単体では従来法の演算量の1/10にできている。このため、目標として上

記の画像強調と影除去の演算量削減目標を一桁以下(10%)となるアルゴリズムを構築する。また、前景分離に関してはそのLSI実現方法についても論じる。演算量を10%に削減し、これに固定小数点演算プロセッサとストリーミング手法を活用して最終目標は100mW以下の実現目標を設定する。

3. 研究の方法

(1) 基本となる前景分離手法

前景分離の手法は既報告した変換領域における混合ガウス背景モデルをベースに本プロジェクトを進める。この方法は画素単位に処理する従来方式より安定性が高く、演算量も1/10ほどに圧縮できる。具体的には、HDTVを対象に毎フレーム前景分離を行うと従来方式では100 GOPS(Giga Operations Per Second)以上の演算量が必要であったが、この手法を用いると10 GOPS以下になる。従来法では画素ごとに入力光の背景確率分布を混合ガウス分布で近似し、この分布領域内に合うと背景とする。つまり、画素数分の混合ガウスモデルを時々刻々更新しつつ、前景か背景かの判断を行っていた。

演算量低減の原理は画面を小さな変換領域に分割し、個々の変換領域で3個の低域変換スペクトルバンドのエネルギーに対応する値を特長量として活用する。このため、画像全体で必要となる混合ガウスモデルの数も激減する。安定性が増す理由も低スペクトル帯域を用いていること、および、変換領域を次第に大きくして各拡大領域の前景判定結果を加味して判断する多重解像度処理を内在させていることに起因している。

この方法は電子情報通信学会英文論文誌 E92-A の 772-778 頁(2009)に詳しい。

(2) 画像強調による前景分離の改善

前景分離を正確に行うためには、種々の画像強調方式のうち、網膜(retina)と脳の皮質の反応(cortex)を組み合わせてシミュレートするレチネックス(Retinex)法が優れた性質を示す。特に脳皮質の刺激反応部分の拡がりガウスフィルタのインパルス応答に対応させ、このフィルタのscale値をこの空間拡がりに対応させるもので、空間的な平均を求めている。入力値と空間平均値との比はコントラストと呼ばれ、これを対数領域で強調する。複数のスケールで画像強調して、個々の強調画像の重み付け平均を取るマルチスケールレチネックス法(以下MSR)と呼んでいる。これに必要な演算量を1/10000にする。ただし、1/1000はよく知られているマルチレート信号処理で実現できる。それ以外で1/10にする。これは、画像のスペクトルの一般的な性質を用いた近似を行って行く。

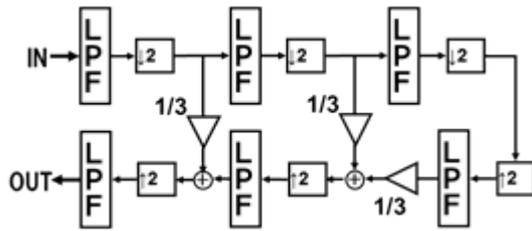


図 1. 提案するマルチレートマルチスケールレチネックスのブロック図。ここで、LPF は低域通過ガウスフィルタ。

(3) 影除去

既存の影除去方式はすべて画素単位の処理方法で、一旦 RGB 表現された画素値を HSV 色空間に変換する。光の強度 (V) と色彩環での角度 (色相 : H) と、鮮やかさ (彩度 : S) で色空間を表す方式である。この空間では影で暗くなった部分でも強度は小さくなるものの色相と彩度は保存されるという原理を用いて影を検出する。RGB から HSV への変換では強度は容易に求められるものの、色相と彩度は円柱座標や円錐座標を用いる関係で演算量が多い。そのため、変換領域前景分離で用いた領域ベースの処理にして検討する。このため、新たに「背景に投影された影の内部ではスペクトル成分がすべて均一に減衰する」という仮定のもとに実行する。前景分離も変換領域であるため、相性がよくこの機能の演算量も 1/10 以下になると考えられる。

シルエット抽出は以上の技術を組み合わせることで完成する。画像強調と影除去は独立して性能評価を行うが、シルエット抽出が完成した時点で、このシルエットを用いた歩容解析でシルエット抽出自体の性能テストも行う。

(4) 低消費電力前景分離 LSI

演算量を削減したアルゴリズムがロボット等で利用可能なものであることの実証実験も重要である。前景分離、画像強調、影除去のすべてを短期間で LSI 化することはできないので、主に前景分離 LSI を FPGA で設計し、FPGA からフル ASIC デザインでの消費電力を見積もる。この結果、50mW 以下の消費電力で実現できる可能性を示す。前景分離 LSI は演算回路を効率よく活用できるプログラマブルプロセッサアーキテクチャとしてまとめ、ストリーミングアプローチで、極力無駄なメモリを使わないように工夫する。

4. 研究成果

(1) 画像強調

研究の方法でも述べた様に、空間平均値を求める演算量削減を実現するブロック図を図 1 に示す。画像強調の入出力は RGB3 成分の信号であるが、図 1 の平均値フィルタの入力は輝度のみとなる。スケールの大きいガウスフィルタは、スケールの小さいガウスフィルタを多段接続して得られること、さらにガウス

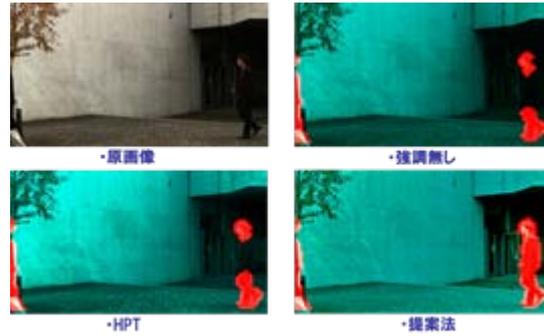


図 2. 画像強調として HPT と提案法を用いた前景分離結果

フィルタは低域通過型フィルタであるので画像を縮小拡大してフィルタ処理を行うマルチレート信号処理を用いること、3 つのスケールを持つフィルタ出力は、このマルチレート処理の中で加算して行っているのが特徴である。図 3 の上段のフィルタ列は画像を順次縦横半分の大きさの画像に間引きしながら処理を行い、下段のフィルタ列は画素間にゼロ画素を挿入して縦横倍の大きさの画像にしてフィルタ処理を行う。3 個のスケールを持つフィルタは上段の途中から下段の途中に係数倍して加算して生成する。係数はこの場合、すべて 1/3 に統一する。上段のフィルタと下段のフィルタは全てガウスフィルタであるため最大スケールのフィルタは上下のフィルタ列で作る。

正確なマルチスケールレチネックスでは平均値を構成する部分で加算してはいけない。コントラストに対応する 3 個の比率を対数領域で加算する必要があるためである。加算の重み付けをすべて 1/3 とすることで図 3 の方式と正しい方式がほぼ等しくなる。この場合、図 1 では平均信号の相加平均を求めており、正しい方式では平均信号の相乗平均を求めているためである。ただし、画像信号は低域にエネルギーが固まっているため、相加平均と相乗平均の差はあまり発生しない。

演算量は入力信号を輝度にしたことで 1/3、最大スケールを画像縮小パスと画像拡大パスで用いることができるので 1/2、さらに、3 個のフィルタを 1 個で代用したので 1/3 となり 1/18 となる。また性能は図 2 に示す暗くなったエントランスの入り口を黒い服で歩いている人の映像に、色相を保持するヒストグラムシフト方式である HPT (Hue Preserving Transform) と提案方式を用いた前景分離結果で、提案法の採用は安定して前景分離ができるのに対し、従来法は胴体が検出できない。また、演算量は従来法と提案法でほぼ同じである。他の細かい結果は [掲載論文 2] に詳しい。

演算量をさらに削減する目的で変換領

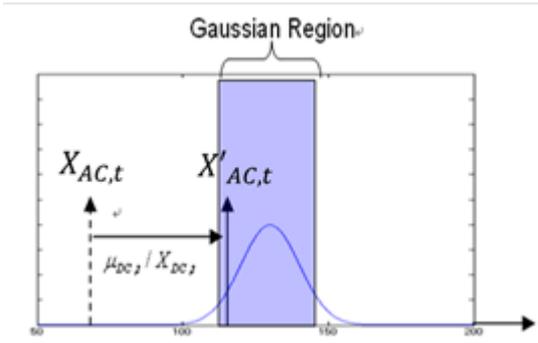


図 3. 影領域の判定法

域アプローチを併用した方式にもチャレンジした。成果は国際会議 ITC-CSCC[学会発表 10]で発表したが、変換領域ごと前景になることも多く、現状では安定した前景分離法には向かないという結果になった。

(2) 影除去

影の定義を従来の HSV 色空間での特徴を利用する従来法から離れて、背景に影が投影されると、すべてのスペクトルが一様に下がるという仮定を用いる。前景分離の特徴量は明るさを表す最低スペクトルと、縦、横方向に伸びた低域スペクトルバンドのスペクトルの絶対値の加算として求めたエネルギー相当量である。よって、背景の影領域ではこれら 3 個の特徴量の減衰度合いは同じになる。

図 3 は提案法の原理を示したものである。背景の混合ガウスモデルは複数のガウス関数で確率密度関数を近似する方法を用いる。図 3 ではその中のあるモードだけに着目した。つまり、1 つのガウス確率密度関数だけを示しており、入力光の特徴量 ($X_{AC,t}$) がこのガウス分布に含まれる場合は背景となる。前景の背景への投影された影部分からの特徴量はこのガウス分布の領域から逸脱し、低い値となる。よって、この影は前景となる。影の候補推定には特徴量は明るさを示す最低スペクトルを用いる。図 2 に示すように、この特影領域の判定法特徴量が含まれていたガウス分布の平均値からの逸脱した比率の逆数を他の特徴量に乗ずる。すると、影の領域ではこれらの特徴量は元のガウス分布に包含される。この基本原理は国際学会 ISIT2010[学会発表 9]で発表した。

しかし、種々の屋外ビデオを用いて評価すると、この影除去法ではシルエットの一部が欠けやすいことに気づいた。原因は輝度信号を用いているため RGB 成分の人間の感覚に合った重み付け加算が行われており、B 成分は R 成分の半分程度に抑えられていることであった。円柱モデルを用いた HSV 色空間の V 成分は RGB 成分の平均値を使う。これまでの画像強調と前景分離の Y 信号入力を V に置き換え、比の計算相当を除算なしで行うことで成功した。この結果は国際学会 ICASSP2011

原画像

シルエット化

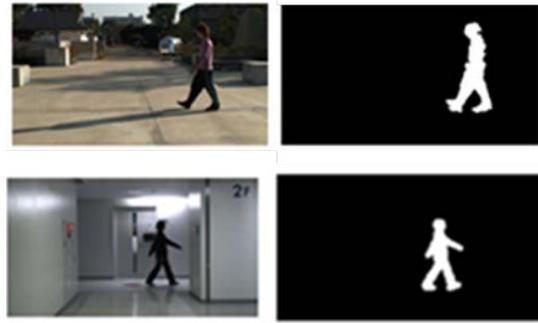


図 4. 屋内外でのシルエット抽出結果

[学会発表 8]で報告した。

この方式も屋内環境では弱点があった。模様のない廊下を無地の T シャツを着た学生が歩く場合に起こる。この場合、変換領域の小さい部分では明るさを示す最低スペクトルの特徴量は明確な前景分離に役立つが、低域スペクトルバンドの特徴量は前景、背景ともほぼゼロである。これで前景部分の低スペクトル領域特徴量を影と考慮して持ち上げてもゼロのまままで前景を背景と誤解して除去するためである。演算量が増加する方向で対応する方式も導出して国際学会 ITC-CSCC2012[学会発表 7]でも発表した。最終的な結論は以下のようになった。

まず、画像強調用の図 1 で示すフィルタ入力は HSV の円錐モデルで使う V を、前景分離は HSV で使う V を用いる。前者の V は RGB 成分の最大のもので V とするため、エネルギーの大きな信号となり、これをさらに画像強調で強化して前景分離に送る。画像強調出力は RGB であるので前景分離では RGB の成分を公平に扱う後者の V が良いということになった。この結果は[掲載論文 1]で詳しく述べている。図 4 は屋内外のシルエット抽出法を示したものである。

以上のようにして抽出したシルエットを用いた歩容解析は国際学会 IWAIT2013[学会発表 3]で報告した。この論文では先に述べた輪郭部の精度向上を狙って輪郭周辺部分だけは 2×2 ブロックを用いた最低スペクトルを特徴量とし、単一ガウスモデルを用いている。この論文では不変アフィン変換を用いた方式でよい結果を出せたことを述べているが、別途 HMM でもテストし、シルエットの面積だけを特徴量として用いている教科書 (Rama Chellappa, etc., "Recognition of Humans and Their Activities Using Video", Morgan and Claypool, 2005) の結果と同程度の品質となった。

(3) 低消費電力前景分離 LSI

本プロジェクトが始まる前から FPGA で

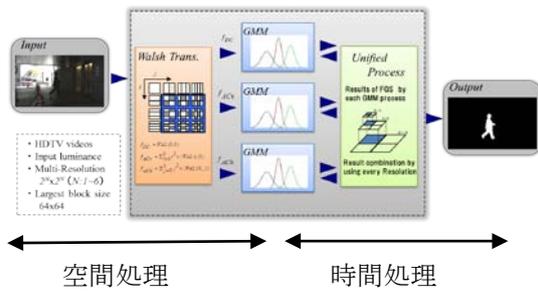


図 5. 空間処理と時間処理の両方をマルチプロセッサ化してチップに入れる。

簡単に信号処理用のアーキテクチャを実現し、評価できるシステムを構築中であった。PC と FPGA ボードを直結し、PC では Mathworks 社のブロック図言語 Simulink でシステムを構築し、その中の複数のブロックを FPGA で置き換える。Simulink を起動させると FPGA で置き換えた部分に到達すると自動的に FPGA 回路が動き出すシステムである。これ等のデバッグ機能もシルエット抽出プロセッサ開発に向けて順次高機能化を図り、国際学会 [学会発表 6, 5, 2] でも積極的に発表した。

前景分離 LSI はアルゴリズムがもっとも早く固まった機能で、前々から LSI 化を狙っていたものである。ストリーミング処理を行うソフトウェアプログラマブルプロセッサ構造を持たせることで大幅なメモリ使用量の削減とアドレス演算をなくすことができるためである。これにより低消費電力化が見込める。

前景分離は、特徴量計算は空間処理であり、この特徴量の混合ガウスモデルを用いた処理は時間処理となる。このため、プロセッサ構造も図 5 に示すように、空間処理用の密結合多重解像度変換マルチプロセッサと、時間処理用のマルチスレッドプロセッサに分割して 1 個のチップ内に収める。この 2 組のプロセッサ間のデータの受け渡しは、1 個の変換領域において 3 個の特徴量が計算されるとすぐに時間処理用マルチスレッドプロセッサが引き取るため、特徴量を伝える 3 個のレジスタだけでよい。

密結合多重解像度プロセッサがストリーミング処理で効率的に実現できる主因は、変換として Walsh 変換を用いたことと、特徴量として変換スペクトルの低域成分だけを用了ことである。多重解像度変換部分は変換領域が 4×4 画素のものから x, y 方向に 2 倍ずつ拡大させる。変換領域が最大となるのは 64×64 画素である。一般的な処理プログラムでは空間処理の変換は同じ大きさの変換領域を持つ部分画像全体を終了させ、次いで時間軸方向の処理を行う。この順序を拡大変換ごとに繰り返す。拡大変化を行うためと時間処理プロセッサへのデータ引き渡しの

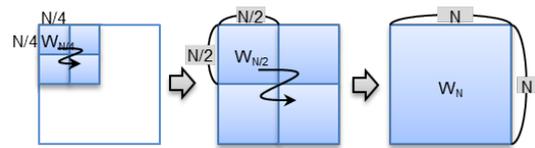


図 6. 高速 Walsh 変換の領域拡大

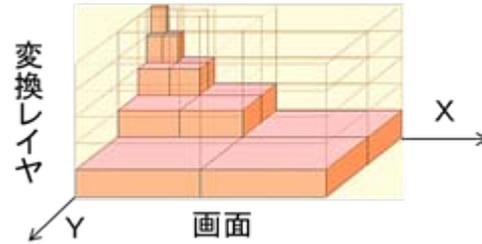


図 7. ストリーミング処理における処理順序のイメージ図

ため変換スペクトルと特徴量を一旦一時メモリに蓄えなければならない。このメモリを排除できるのがストリーミング法である。

まず、Walsh 変換を用いると多重解像度に対応する変換領域拡大は図 6 の高速 Walsh 変換と同じ方法で実現できる。つまり 4 つの隣接する小さな正方領域スペクトルから 4 倍の変換領域スペクトルが生成できる。このような性質はフーリエ変換では実現できない。変換領域が入れ子になったものから拡大領域のスペクトルを求めるためである。

特徴量を変換領域の低スペクトルのみで計算できる利点は、高いスペクトル成分は計算しなくてよいことと、拡大領域の低スペクトルは元の領域の低スペクトルのみで生成できることにある。

これら 2 つの利点と図 6 の性質から、ピクセルフロー法を導いた。これは一度画素データが読み出され変換されると、そのスペクトルを一時レジスタにのみ蓄えておき、隣接する 4 個の同一領域サイズの変換が終われば、直ちに拡大領域の変換処理に移る方法である。残りの場所にある変換処理は新たに必要にならない限り実行しない。つまり、一度読み出したデータは可能な限りレジスタファイルなどに格納するだけで処理する。このことでスペクトル格納用のメモリを排除できる。図 7 はこの実行順序を図示したもので、時間的に早く処理される領域が上になるように描いている。ただし、レジスタ総数は 64 個必要になるが、16 ビット固定小数点演算で行うため一般の組み込み RISC プロセッサにおける 32 ビットレジスタ 32 個と同程度になる。

一方の時間処理マルチスレッドプロセッサは次ページの図 8 に示すように、3 個の特徴量処理に一個ずつ同一構造のプロセッサ

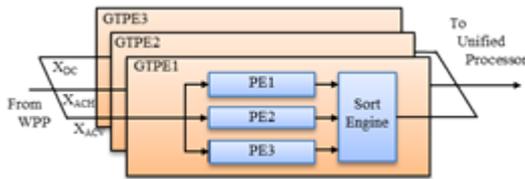


図 8. 時間処理マルチスレッドプロセッサ

サを準備し、この同一構造のプロセッサは混合ガウスモデルで用いるガウス分布数だけのパイプライン演算器を持ったプロセッシングエレメントを設ける方法で実現した。

以上のアーキテクチャに基づいてハードウェア記述言語でシステムを構成した。実際にはホストプロセッサからのデータ転送が必要となるものの、Xilinx 社の Virtex-4 シリーズで最も小さい XC4VLX15 を用いても 50%未満の占有率しかなく、ハードウェア記述言語から消費電力を見積もる X-power ソフトウェアで推定しても 150mW である。フル ASIC 設計をすると消費電力は 20%前後に減少するため、30mW 相当となり、ロボットやモバイル端末向けに目標としていた消費電力の目標を十分に達成した。この成果は国際学会 ICASSP2012[学会発表 4]で発表した。また、多重解像度蜜結合プロセッサの詳細設計は電子情報処理学会 SIP 研究会[学会発表 1]で報告し、マルチスレッドプロセッサは開発システムを含めて国際会議 IWAIT2013[学会発表 2]で報告した。

5. 主な発表論文等

- [1]Toshiaki Siota, Kazuki Nakagami, Takao Nishitani, “Transform Domain Shadow Removal for Foreground Silhouette”, IEICE Trans. Fund, 査読あり, Vol. E96-A, No. 3, pp. 667-674, March 2013.
- [2]Takeshi Okuno and Takao Nishitani, “Efficient Image Enhancement Algorithm Using Multi-Rate Image Processing”, IEICE Trans. Fund., 査読あり, Vol. E93-A, No. 5, pp. 958-965, May, 2010.

[学会発表] (計 10 件)

- [1]梶本、八木、西谷、「実時間前景分離FPGAの実装に関する一検討」、電子情報通信学会信号処理研究会資料、査読なし、SIP2012-83, 2013年1月31日, 呉.
- [2]R. Yagi and T. Nishitani, “Implementation of Programmable Processors by using SW/HW Co-Simulator”, 電子情報通信学会等, Proc. IWAIT2013, 査読あり, p. 790, Jan. 6, 2013, 名古屋.
- [3]T. Shiota and T. Nishitani, “Precise and Stable Silhouette Extraction”, 電子情報通信学会等, Proc. IWAIT2013, pp. 1074-1079, Jan. 6, 2013, 名古屋.

- [4]R. Yagi, T. Kajimoto and T. Nishitani, “GMM Foreground Segmentation Processor based on Address Free, Pixel Flow Streams”, IEEE SPS, Proc. ICASSP2012, 査読あり pp. 1653-1656, March 25, 2012, 京都

- [5]Y. Takakanishi and T. Nishitani, “Application on the Hardware/Software Co-Simulator; Implementation of Multi-stage, Multi-rate 2D Filter”, IEEE 等, 査読あり Proc. SASIMI2012, pp. 168-173, March 8, 2012, 別府.

- [6] Y. Takakanishi, K. Mori, T. Nishitani and Y. Nakamura, “Hardware/Software Co-Simulator for ASIC DSP Chips”, IEEE CAS, 査読あり, Proc. MWSCAS2010, pp. 809-812, 2011, Aug. 7, 2011, シェアトル.

- [7]T. Shiota and T. Nishitani, “Efficient and All-Round Shadow Removal Method in Transform Domain”, 電子情報通信学会等, 査読あり, Proc. ITC-CSCC2011, pp. 1083-1086, June, 19, 2011, Gyeong 韓国.

- [8]K. Nakagami, T. Shiota and T. Nishitani, “Low Complexity Shadow Removal on Foreground Segmentation”, IEEE SPS, 査読あり, Proc. ICASSP2011, pp. 1081-1084 May 22, 2011, プラハ.

- [9]K. Nakagami and T. Nishitani, “The Study on Shadow Removal on Transform Domain GMM Foreground Segmentation”, IEEE, 査読あり, Proc. ISCIT2010, pp. 867-872, Oct. 26, 2010, 東京.

- [10]Y. Akasaka and T. Nishitani, “High Precision Detection on Computationally Efficient Silhouette Extraction”, 電子情報通信学会等、査読あり、Proc. TTC-CSCC2010, pp. 288-291, July 4, 2010, パタヤ, タイ.

[その他]

ホームページ等

現時点で退職に伴い、大学のホームページ閉鎖

6. 研究組織

(1) 研究代表者

西谷 隆夫 (NISHITANI TAKAO)

首都大学東京システムデザイン研究科・教授

研究者番号：00389206