

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 24 年 6 月 11 日現在

機関番号：21401

研究種目：若手研究(B)

研究期間：2009～2011

課題番号：21700257

研究課題名（和文） 適応的学習による時空間自動ラベリングの研究

研究課題名（英文） Study on Automatic Spatiotemporal Labeling Using Adaptive Learning

研究代表者

間所 洋和 (MADOKORO HIROKAZU)

秋田県立大学・システム科学技術学部・助教

研究者番号：10373218

研究成果の概要（和文）：

本研究では、適応共鳴理論と自己組織化マップを用いてカテゴリマップを自動で生成する手法を開発した。カテゴリ数が未知な分類問題である表情分類を対象として、本研究で独自に作成した表情画像データセットを用いて評価した。本データセットは、10名の被験者に対して最大20週間の画像列から構成される。評価実験では、表情の空間的な分布と時間的な関係が同時に表現できた。また、ロボットビジョンにおける一般物体認識に応用し、提案手法の拡張性を示した。

研究成果の概要（英文）：

The aim of this study was to develop a method to create category maps using Adaptive Resonance Theory (ART) and Self-Organizing Maps (SOM). As a classification problem that is unknown the number of categories, we evaluated our method to classify of facial expression images. We developed an original dataset of facial expression images consisted of 10 subjects during twenty weeks in one week interval. For the evaluation experiments, our method was able to describe the relations between spatial distribution of and time-series features on dynamic diversity of facial expressions. Moreover, we applied our method to generic object recognition on robot vision to show the extensivity of category maps.

交付決定額

(金額単位：円)

| | 直接経費 | 間接経費 | 合計 |
|--------|-----------|---------|-----------|
| 2009年度 | 2,000,000 | 600,000 | 2,600,000 |
| 2010年度 | 800,000 | 240,000 | 1,040,000 |
| 2011年度 | 500,000 | 150,000 | 650,000 |
| 年度 | | | |
| 年度 | | | |
| 総計 | 3,300,000 | 990,000 | 4,290,000 |

研究分野：総合領域

科研費の分科・細目：情報学・感性情報学・ソフトコンピューティング

キーワード：ソフトコンピューティング、適応共鳴理論、自己組織化マップ

1. 研究開始当初の背景

創造性や知覚、思考、記憶などの脳の高次機能を支える大脳皮質は、教師なし学習に特化した振る舞いを示唆することが、Doyaらに

より神経科学的に示されている。Kohonenによって提案された自己組織化マップ (Self-Organizing Maps: SOM) は、教師なし学習の代表的なニューラルネットワーク

であり、競合と近傍の概念により入力データの位相構造を保存しつつ低次元空間へと自己写像する。特に、ベクトル圧縮による結合荷重の可視化特性に優れていることから、画像処理やコンピュータビジョンをはじめとする様々な問題に応用されている。

学習後の SOM は、Kohonen 層のユニットに形成されるカテゴリを同定するために、ラベルを与える必要がある。SOM は高い写像能力を有するが、その分類性能は、写像結果に対するラベリングに大きく影響される。Nilsen が提案した対抗伝搬ネットワーク (Counter Propagation Network: CPN) は、SOM にラベリングを担う Grossberg 層が追加された自動ラベリングネットワークである。しかしながら、CPN は教師あり学習であるため、ラベリングには教師信号が必要となる。

教師なし学習におけるラベリングは極めてプリミティブな作業であり、特に、マッピング層のユニット数が多い場合には多大な負荷を要する。また、ラベリングを行うための明確な指標も確立されておらず、多くの場合、設計者や実験者の主観的な判断による手作業に頼っているのが現状である。更に、教師なし学習では、カテゴリ数が既知か未知かで対象問題の取り扱いが大きく異なる。教師なし学習で扱う多くの問題は、カテゴリ数が未知であるにもかかわらず、これまでの研究では、その評価方法の困難さから、カテゴリ数が既知という前提で手法の評価が行われていた。

2. 研究の目的

本研究では、適応共鳴理論 (Adaptive Resonance Theory: ART) を用いて、SOM のカテゴリマップを自動でラベリングする手法について研究する。本研究で提案する手法は、安定性と可塑性を併せ持つ ART と位相関係を自己写像する SOM が時間関係と空間関係をそれぞれ学習することにより、カテゴリマップに時空間関係を与える。カテゴリ数が未知の分類問題である表情分類を対象として、公開表情画像データベースと本研究で新たに作成する表情画像データベースを用いて提案手法の性能を評価する。

表情の空間的かつ時間的なカテゴリ変化を静的多様性及び動的多様性として扱うことにより、前者では基本 6 表情に対して 95% 以上の識別率、後者では個々人で異なる表情のカテゴリ数の抽出と各被験者の心理学的主観評価で 90% 以上の一致率を目標とする。更に、ニューラルネットワークの本来の処理体系である並列分散処理とそれに伴う高速化の実現を目指して、FPGA (Field Programmable Gate Arrays) 上にハードウェア実装する。

3. 研究の方法

顔という対象の形状が人物それぞれで異なるように、表情表出の仕方、例えばある感情をどの程度の大きさの顔面変形として表情に表出するかについては個人差がある。このため、法政大学の赤松は、個々人の表情表出の特性に応じて修正していく適応的な学習メカニズムが必要と述べている。したがって、本研究では、被験者が意図的に表出した表情を対象として、時間軸方向への圧縮による正規化と表情表出における位相変化を抽出するために SOM を用いて表情パターンの分類を行う。更に、SOM により分類した表情画像を、安定性と可塑性を併せ持った適応的学習アルゴリズムである Fuzzy ART を用いて再分類する。SOM は、予め決められた写像空間の中で相対的にカテゴリ分類を行うが、Fuzzy ART は、ビジランスパラメータで制御された一定の粒度のもとでカテゴリ分類を行うため、長期間に及ぶ時系列データに対しても、同じ基準で分類することができる。

(1) 前処理

本研究では、AU (Action Units) による特徴点ベースの手法ではなく、画像全体から特徴を表現する見え方ベースの手法を用いる。表情表出に関する局所的な特徴量の変化は、見え方ベースよりも特徴点ベースの方がより細かな記述が可能である。一方、特徴点ベースは、特徴点を検出し追跡するための処理に多くの計算負荷を要するとともに、大量のサンプルに対して自動的に処理する際に、精度及び安定性において課題が残っている。

本手法では、人間の視覚特性に基づき、Gabor Wavelets で特徴表現された画像を用いた見え方ベースの手法を採用する。特に、本研究で対象とした処理は、無表情から単一の表情を表出した際のパターンの変化から覚醒度を抽出することであり、変化部位の出現範囲が限定されるため、AU により特徴点を追跡しなくても、その部位の変化が Gabor Wavelets 変換後の特徴空間上に出現していると考えられる。最後に、情報量の圧縮とノイズの低減を目的として、粗視化処理を行う。

本手法では、顔画像を取得する際に顔の構成パーツ (眉、目、鼻、口、頬、顎) を含む領域を手動でテンプレートの初期位置として設定していることや、顔画像撮影中におこる被験者の体動により微少のずれが生じてしまうため、粗視化処理によってこのような影響を緩和している。なお、粗視化ウィンドは、予備実験の結果から 10×10 画素とした。したがって、 80×90 画素の原画像は、 8×9 画素に圧縮される。

(2) SOM による表情パターンの分類

1 セット 200 枚から構成される時系列画像から、覚醒度に応じた分類を実現するために、対象画像セットを一定の範囲内で正規化す

る必要がある。本手法では、教師なし学習により競合と近傍の概念に基づき任意のユニットに対応付くカテゴリ分類を行う SOM を用いる。

SOM のネットワーク構成は、入力層とマップ層の 2 層から構成される。マップ層の各ユニットは入力層の全ユニットと結合しており、結合荷重を保持する。入力データに対して、最も類似する結合荷重を保持するユニットが発火し、そのユニットと近傍を形成するユニットが、入力データに近づくように結合荷重を更新することにより、SOM の学習が行われる。入力データ間の類似性が、1 次元または 2 次元配列上で発火するユニットの近さに反映されるトポロジ保存性という特徴を持つ。学習が進むにつれて、類似した特徴を持つデータは近くに、そうでないものは離れた位置に配置されるようなマップが形成される。

SOM は各データセットに対して個別に適用する。予備実験の結果から、マップ層は 15 ユニット、学習回数は 1000 回とした。学習時は各ユニットに偏りが生じないように、入力画像は 200 枚の中からランダムに選択した。学習後の SOM に表情画像を再入力することにより、各ユニットに対応付ける。

(3) Fuzzy ART による表情パターンの統合

SOM によるカテゴリ分類は、予め定められた写像空間、すなわちマップ層のユニット数の中で分類が行われるため、分類結果は相対的となる。一方、縦断的データセットとして長期間に渡るデータに対して、同じ基準、すなわち、一定の粒度で分類を行う必要がある。そこで本手法では、SOM で分類されたカテゴリを代表する結合荷重を学習データとして Fuzzy ART で再分類することにより、表情パターンを統合する。

ART は、過去の記憶を保持しつつ新しいパターンの学習が可能で、安定性と可塑性を併せ持つ追加学習型の教師なしニューラルネットワークである。ART には様々な種類が存在するが、本研究ではアナログ入力に対応した Fuzzy ART を使用する。Fuzzy ART のネットワーク構成は、入力層の F0、特徴表現の F1、及びカテゴリ表現の F2 から構成されている。

Fuzzy ART では、F2 のユニットに SOM の結合荷重が対応付く。このとき、類似した特徴量はカテゴリとして統合される。最後に、各カテゴリの平均顔画像を、無表情を基準にして覚醒していくパターン順に並べ替えられる。ここで、対応付く画像の枚数が最も多いカテゴリを無表情のカテゴリとする。これは、本データセットには無表情となる画像を多く含むという暗黙知を用いている。表情空間チャートでは、中心部が覚醒度 0 の無表情となり、三角形が外側に広がるにつれて覚醒

度の大きな値を示すカテゴリを配置する。

(4) 表情空間チャートの生成

SOM で分類され Fuzzy ART で統合された表情カテゴリは、無表情を基準として覚醒度順に並べ替えられる。当該データセットには、無表情画像が最も多く含まれている。したがって、画像枚数が最大のカテゴリを無表情のカテゴリとする。覚醒度の並びは、無表情を基準に相関の低いカテゴリの順とした。

表情空間チャートの各軸には、覚醒度の最大値が対応付けられる。表情空間チャートの中心部は、覚醒度 0 (無表情) となる。覚醒度が大きくなるにつれて、三角形の外側に表情カテゴリが配置される。

4. 研究成果

(1) データセットの構築

表情認識用の顔画像データセットは、大学や研究機関において様々な撮像条件で作成したものが、インターネット等を通じて一般に公開されており、特に性能比較のために活用されている。静止画像に関しては、表情認知の心理実験で視覚刺激とする目的で収集された Ekman&Friesen のデータセットが有名である。動画画像では、Cohn-Kanade や Ekman-Hager などのデータセットが有名であり、性能評価及び比較実験用として幅広く利用されている。また、近年では、Panticらが公開している MMI Facial Expression Database が著名である。これらのデータセットは、被験者数としては十分な人数が確保された横断的なデータセットであるものの、各被験者に対して 1 回のみ撮影に留まっており、同一被験者に対して長期的に表情を撮影したデータセットは存在しない。

表情の表出は、撮影時点での被験者の心理状態や文脈、状況などによって、表出度合いやパターンが異なる。本研究では、個人固有な表情空間モデルの作成とストレスレベルの推定を行うために、表情の撮影期間を長期に設定した。また、表情の表出パターンは個人差の影響を強く受けるものの、既存の手法の大半は、複数人に対して単一のモデルを構築し、識別や認識を行っていた。本研究では表情空間チャートを形成する覚醒度の抽出及びストレスレベルを推定するための推定器の構築は、個人毎のモデルを作成する。個人毎のモデルを作成するには、1 回の撮影だけではなく、複数回のデータが必要となるため、本研究では長期間に渡る縦断的的表情画像データセットを構築した。具体的には、10 名の被験者を対象として、数カ月をわたり表情画像を取得した。被験者の内訳は、女子大学生 5 名 (A, B, C, D=19 歳, E=21 歳)、男子大学生 5 名 (F, J=19 歳, G, H, I=22 歳) の合計 10 名である。撮影期間は被験者によって異なるが 7~20 週間 (1 週間間隔) とし

た。

日常生活で頻繁に目にする表情は、単一表情より複合表情が多い。よって、単一表情のみではストレスレベル推定への影響が生じると考えられる。しかしながら、複合表情を表情画像データセットとして大量に取得することは、被験者への訓練、理解、心理的負担を考えると、非常に困難な課題である。また、複合表情は、複合される各表情の比率を定量的に把握することが難しい。したがって、本研究では単一表情のみを対象とした。

本研究の表情撮影は長期間に渡るため、実験に協力してくれる被験者の負荷軽減を考慮して、基本6表情から撮影対象とする表情を3表情まで絞り込むことにした。始めに「嫌悪」に関して、Ekman は表情表出における文化的要因として、日本人は「嫌悪」を感じているにもかかわらず笑顔を表出するという Japanese smile について言及している。このことから、日本人にとっては「嫌悪」の表出は難しいと判断し、撮影対象外とした。次に「恐れ」の表情に関しては、日常生活でそのような状況になることが稀であり、表出の仕方がよく分からないといった意見が多かったことから撮影対象外とした。最後に「驚き」に関しては、表出への抵抗感は少なかったが、特に男性の被験者を中心に照れの感情が入る傾向にあった。このため、笑顔と混在した複合表情になる画像が多数見受けられたことから、撮影対象外とした。以上の理由を鑑みつつ本研究では、表情表出の容易性と Russell の円環モデルにおける各象限での布置を考慮して、撮影対象を「喜び」「怒り」「悲しみ」の3表情とした。

撮影環境は、通常室内（日常一般的と考えられる蛍光灯による照明下）とし、被験者の頭部がフレーム中に含まれる状態で正面顔画像を撮影した。あらかじめ被験者には、頭部をあまり動かさないで表情を表出するように指示して撮影したため、一定の範囲内に顔領域が収まっているが、微小な変動に対しては、Haar-like 特徴と Boosting を用いた顔領域のトラッキング手法を用いて補正した。撮影の容易性と今後の応用展開を考慮して、カメラは市販の USB カメラ (Logicool 社製 QcamOrbit) を用いた。被験者が意図的に表出した表情画像列（無表情と表情表出の繰り返し）を1セットとした。撮影時間は20秒間として、被験者にはこの間に表情表出を3回程度繰り返すよう指示した。カメラのサンプリングレートは毎秒10フレームに設定したため、1セット200フレームの画像列から構成される。

撮影条件として被験者には、以下の内容を事前に説明した。

- ・撮影中はカメラに向かって表情を表出する。
- ・頭部をあまり動かさないようにする。

・20秒間の撮影時間内に表情表出を3回繰り返す。

- ・各表情とも最大まで表出する。
 - ・表情表出後は必ず無表情に戻ってから表出する。
 - ・緊張せずリラックスした気分で表出する。
- なお、被験者には、FACS によるトレーニングは行っていない。また、撮影した表情画像を学術研究以外では使用しないという誓約書への署名のみを実施し、本研究の具体的な内容やストレスレベルの推定については説明していない。

同時に本研究では、心理的ストレス反応測定尺度として、鈴木らが提案した SRS-18 (Stress Response Scale - 18) を用いた。SRS-18 は、日常生活で経験する心理的ストレス反応を短時間で簡易かつ多面的に測定することができる質問紙から構成されている。心理的ストレス反応としては、日常的に体験する各種ストレスによって引き起こされる、憂うつ・不安や怒り（情動的反応）、無気力や集中困難（認知的反応）、仕事の能率の低下（行動的反応）などがある。

測定内容としては、抑うつ・不安、不機嫌・怒り、無気力の3因子に対するストレス反応が対象となる。18項目の質問に対して、回答は「全くちがう」から「その通りだ」の4件法であり、それぞれに0~3ポイントの得点が与えられる。得点範囲は0~54ポイントであり、得点が高いほどストレスが高いことを意味する。更に、この得点から4段階の評定値（レベル1: 弱い、レベル2: 普通、レベル3: やや高い、レベル4: 高い）に分類される。本実験では、被験者10名を対象として、表情撮影（7週間~20週間）を行うと同時に、SRS-18 を用いて心理的ストレスを測定した。

SRS-18 への記入は表情撮影前に行い、表情の表出に影響を与えないようにするために、得点は被験者に提示していない。なお、被験者によっては表情撮影による一過性ストレスの発生や本撮影を続けることによって生じる長期的ストレスの影響が考えられるが、SRS-18 は数日間程度の慢性的ストレスを計測対象としているため、本実験による一過性及び長期的ストレスによる影響は少ないと考えている。

なお、SRS-18 は主観評価であるため、被験者によっては主観のブレなどが推定精度に影響を与えることが考えられる。本研究では、客観的なストレス値を得るために、唾液アミラーゼ活性によりストレスを測定するニプロ社の唾液アミラーゼモニタの使用について検討した。唾液アミラーゼモニタによるストレス評価の有用性は、柏野らによっても示されているが、測定前の数時間は食事を控えることや、測定時に唾液を採取するための専

用のチップを口の中に 30 秒程度入れなければならないなど、データ取得における被験者への負荷が大きい。また、唾液アミラーゼモニタは、一過性のストレスに対して敏感に反応するため、測定データのばらつきが大きいことから、本研究での使用を見送った。なお、SRS-18 は 18 項目の質問のみから構成されているため、短時間でシートの記入ができるとともに、健常者が日常的に接する事象に関する質問項目が多く、更に、心身の諸反応を多面的に測定できることから、表情との併用に最適かつ有用なチェックシートと考えられる。

(2) 表情カテゴリの生成結果

本手法では、各被験者の各週の各表情に対して個別に SOM を構築し学習する。入力層は、粗視化後の入力画像のサイズと同じ 72 ユニットとした。マップ層のユニット数に決定方法については、全画像データセットを対象とした予備実験において、マップ層のユニット数を 9 ユニットから 21 ユニットまで変化させることにより SOM の写像結果を比較検討した。その結果、15 ユニット以上の場合において学習データが対応付かない未発火ユニットが発生したことから、SOM のネットワーク構成が最小でかつ表現能力が最大となる 15 ユニットとした。学習回数は 500 回とした。

SOM の学習結果でありかつカテゴリの代表ベクトルとなる結合荷重を学習データとして Fuzzy ART の学習を行い、覚醒度を決定するためのカテゴリを一定の粒度で統合する。Fuzzy ART のネットワーク構成は、SOM と同様に F1 層は 72 ユニット (8 × 9 次元)、F2 層は 15 ユニットとした。Fuzzy ART の動作を特徴付けるビジランスパラメータ、選択パラメータ、学習率パラメータは、それぞれ、0.90、0.75、0.50 に設定した。これらのパラメータの中でも、分類粒度を決定するビジランスパラメータは、Fuzzy ART のメタパラメータとして機能するため、予備実験により表情変化に伴う位相構造がカテゴリとして反映されていると判断される結果から主観的に決定した。しかしながら、このようなパラメータの決定は、客観性に欠け、実験者に対する負荷も大きいことから、進化的学習やメタ学習によるパラメータの最適化手法等の導入を検討している。

(3) 表情空間の定量化結果

覚醒度として対応付けるために、SOM で分類し Fuzzy ART で統合したカテゴリを抽出する本実験条件では無表情の画像が多くなるため、対応付く画像枚数が最大となるカテゴリを無表情カテゴリとする。この基準カテゴリからカテゴリ間の相関を算出し、相関が小さいカテゴリから順に並び替えることによ

り、各カテゴリと覚醒度が対応付けられる。

ここで、覚醒度が表情表出に応じて対応付けられていることを検証するために、表情表出に伴う覚醒度の時間変化について評価した。各表情の覚醒度の時間変化を図 1 に示す。横軸は画像のフレーム数であり、各画像セットとも 200 フレームから構成されている。縦軸は覚醒度を示す。事前に原画像から表情の表出が開始されているフレーム位置と表出表情から無表情に戻る位置を同定し、図中に破線の縦線で示している。被験者は撮影 1 セット (20 秒間) で 3~4 回の表情表出を繰り返しており、このデータセット (被験者 A, 9 週目) では、「笑い」が 3 回、「怒り」と「悲しみ」が 4 回ずつ表出している。表情表出の開始と終了のタイミングが覚醒度の変化として表現されている。また、若干の揺らぎが含まれるものの、表情表出に対応付くように覚醒度が変化している。

表情空間チャートの例を、図 2 に示す。同一人物であっても、週単位で表情パターンが変化していることが本チャートを用いることにより観察できる。このような変化は、心理面の影響を受けていると考えられるため、本結果は SRS-18 を用いた心理的ストレステストの結果と併せて考察する。

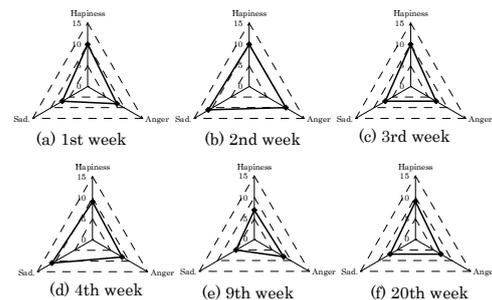


図 1. 表情空間チャート (被験者 A)

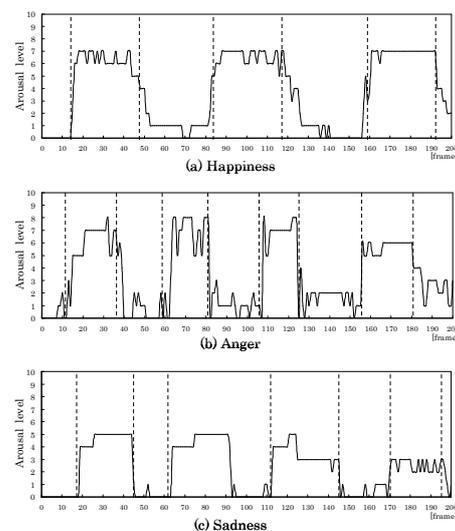


図 2. 覚醒度の時間変化

(4) ロボットビジョンへの応用

SOM と ART を組み合わせることにより、入力特徴量の時間的かつ空間的関係性をカテゴリマップ可視化する提案手法を、ロボットビジョンにおける一般物体認識に応用した。

一般物体認識においてベンチマークテスト用の標準データセットとして広く普及している Caltech256 を用いて、提案手法の基本的特性を評価した。本実験では、256 種類のカテゴリから、データ数が多い上位 20 カテゴリを評価対象とした。各カテゴリからランダムに 5 枚の画像を取り出し、合計 100 枚の画像を用いて学習を行った。

CPN によるカテゴリマップの生成結果を図 3 に示す。カテゴリマップ上には、ユニットに写像された代表画像を表示している。5 カテゴリにおける分類では、同一カテゴリの画像は隣接したユニットに写像された。また、独立した領域が形成された。20 カテゴリにおける分類では、11 カテゴリの画像が隣接したユニットに写像された。また、ART-2 では複数のラベルに分割されていたカテゴリが、CPN では隣接したユニットに統合された。

続いて、5, 10, 20 カテゴリにおける学習とテストの認識率を図 4 に示す。ここでは、OC-SVM を適用した場合と未適用の場合を比較することにより、特徴点選択の有用性を検証した。OC-SVM を適用した場合は、学習データが 96%, 94%, 81%, テストデータが 76%, 42%, 45% となった。一方、未適用の場合は、学習データが 84%, 70%, 64%, テストデータが 76%, 30%, 38% であった。以上の結果から、学習データでは 18%, テストデータでは 6% の精度向上がみられ、OC-SVM による特徴点選択の有用性が示された。

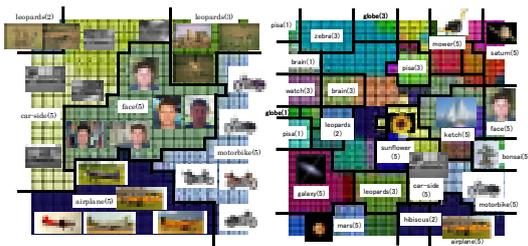


図 3. カテゴリマップ

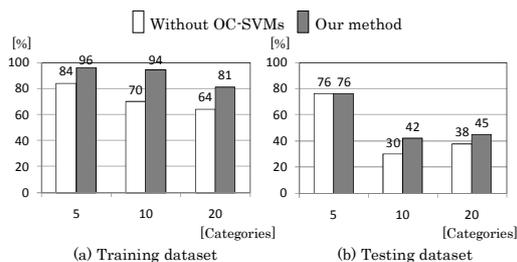


図 4. 認識率の比較結果

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 2 件)

① 間所洋和, 佐藤和人, 門脇さくら, “表情の時系列変化を可視化する表情空間チャート,” 知能と情報 (日本知能情報ファジィ学会誌), Vol. 23, No. 2, pp. 157-169, Apr. 2011.

② M. Tsukada, Y. Utsumi, H. Madokoro, and K. Sato, “Unsupervised Feature Selection and Category Classification for a Vision-Based Mobile Robot,” IEICE Trans. Inf. & Sys., Vol. E94-D, No. 1, pp. 127-136, Jan. 2011.

[学会発表] (計 32 件)

① H. Madokoro, Y. Utsumi, and K. Sato, “Unsupervised Scene Classification Based on Context of Features for a Mobile Robot,” Proc. 15th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES), Part I, pp. 446-455, Sep. 2011.

② H. Madokoro, M. Tsukada and K. Sato, “Unsupervised Feature Selection and Category Formation for Generic Object Recognition,” Proc. 14th International Conference on Computer Analysis of Images and Patterns (CAIP), pp. 427-434, Aug. 2011.

③ H. Madokoro, M. Tsukada, and K. Sato, “Unsupervised Feature Selection and Category Formation for Mobile Robot Vision,” Proc. IEEE International Joint Conference on Neural Networks (IJCNN), pp. 320-327, Jul. 2011.

他

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 0 件)

[その他]

ホームページ等

6. 研究組織

(1) 研究代表者

間所 洋和 (MADOKORO HIROKAZU)

秋田県立大学・システム科学技術学部・助教

研究者番号: 10373218