

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 04 月 27 日現在

機関番号：12612

研究種目：基盤研究(C)

研究期間：2009～2011

課題番号：21500213

研究課題名（和文）

識別モデルと平均場モデルの融合による機械学習と画像認識への応用

研究課題名（英文）Machine learning via fusion of discriminative and mean field models and its application to image recognition

研究代表者

高橋 治久 (TAKAHASHI HARUHISA)

電気通信大学・大学院情報理工学研究科・教授

研究者番号：90135418

研究成果の概要（和文）：

機械学習には、データの確率分布による生成モデルとそれを考慮しない識別モデルがある。本研究では、この二つのモデルを組み合わせることにより、識別性能を上げる事を目的とし二つの主な成果を得た。一つは、識別ランダムフィールドと呼ばれ、生成モデルに識別機能を直接設けたモデルである。ビデオ画像識別に応用した。もう一つは、生成モデルにカーネルを導入したカーネルランダムフィールドであり、画像分類に応用した。

研究成果の概要（英文）：

In machine learning, we have two typical models, i.e., generative and discriminative models. We aimed to combine both of these models in order to obtain more elaborated machine learning models. Two main results are obtained as follows. Firstly, we designed a special discriminative random field and applied it to high performance video image classification. Secondly, we modeled a kernel random field which is constructed as random field with kernels, and successively applied to scene classification.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2009 年度	1,100,000	330,000	1,430,000
2010 年度	1,000,000	300,000	1,400,000
2011 年度	1,200,000	360,000	1,560,000
総計	3,300,000	990,000	4,290,000

研究分野：情報通信(共通基礎研究)

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：機械学習，画像識別，識別モデル，生成モデル，ランダムフィールド，カーネル

1. 研究開始当初の背景

ビデオ画像は、生活、産業、環境など至る所で撮影されその量は日々膨大なものとなっている。これら動画画像から必要な情報を取り出す動画画像理解が近年の課題であるが、その技術は現在発展途上である。動画画像理解の基本は、対象の検出・識別と追跡であるが、

受動的な情報取得にとどまらず、技術の先には、複数の対象の動きから、予測を行い事故を未然に防いだりする能動的対応が期待され、基本技術の飛躍的進展が望まれている。

2. 研究の目的

(動)画像理解では、対象を確率的に扱うた

め、中核となる最も重要な処理技術は機械学習である。ところが、機械学習による処理は計算負荷が大きく画像理解のネックとなっており、その飛躍的発展が、今切望されている。

効率的で的確な動画理解を実現して行くにあたり、機械学習に要求されることは、
 (1) 小さいサンプルサイズで、
 (2) 高速な学習が出来、
 (3) 運用に際しては計算負荷が小さく高速処理が出来、
 (4) 認識性能が高く、
 (5) 識別・生成またはその融合された機能での学習を状況に応じて適用できることである。

現在、画像処理における機械学習において、これらの要請の(1)(2)(3)あるいは(2)(3)(4)が実現できていない。また(5)を備えたものはごく限られている。これらのことが動画処理の性能向上・実用化に向けて大きな壁となっている。本研究では、H19年度で修了した基盤研究(C)の研究遂行の最終年度に得られた、平均場・識別モデル融合のアイデアをもとに上記(1)から(5)すべての条件を満たす、機械学習を確立し(動)画像理解への応用によってその革新的進歩を計ることを目的とする。

3. 研究の方法

平均場近似により構成した特徴空間に対し、特徴の線形分離による識別モデルを組み合わせることで、生成モデルと識別モデルの融合モデルを提案する。このモデルの性能を引き出す検討をした後、適用範囲を広げる研究を行うと共に、応用に合わせた提案手法の効率的な適用の仕方について具体的な検討を集中的に行いこの手法を発展させる。この手法を主として画像処理に応用する。

4. 研究成果

平均場近似法を適用するためには、カーネル関数が、周辺分布平均をとったとき、変数の周辺分布平均のカーネル関数に等しくなる性質が要求される。この性質を満たすためには、カーネルが、変数毎に線形な成分を持つ非線形写像の内積として表現されれば十分である。この性質を満たすのは、一般的多項式カーネルではなく、基本対称式の各項を成分に持つような特徴写像に分解される多項式カーネルを定義すればよい。この目的で以下の漸化式により $O(n)$ で計算できる対称多項式カーネルを提案した。

$$S_k = \sum_{i=[k/2]_+ - 1}^{n-[k/2]_+ + 1} \hat{S}_{[k/2]_-}^{i-1} S_{[k/2]_+}^i$$

$$K(x, z) = \sum_{i=0}^d S_i(y), \quad y = (y_1, \dots, y_n)^t$$

このカーネルを用いるカーネルランダムフィールドは以下のようになる。

$$P_\mu(x) = \frac{1}{Z} \exp \left(\sum_{\ell=1}^m \mu_\ell K(\xi_\ell, x) \right)$$

ここでポテンシャル関数が以下のように定義される。

$$V_\mu(x) = - \sum_{\ell=1}^m \mu_\ell K(\xi_\ell, x).$$

このままでは計算は膨大であるため平均場をとり周辺確率を求める式が以下で与えられる。

$$q_i^k = \frac{\exp \left(- \frac{\partial E\{V_\mu(x)\}}{\partial q_i^k} \right)}{1 + \sum_{k'=1}^{r-1} \exp \left(- \frac{\partial E\{V_\mu(x)\}}{\partial q_i^{k'}} \right)}$$

但し E は平均を表し

$$E\{V_\mu(x)\} = \sum_{\ell=1}^m \frac{\partial E_r\{V_\mu(x)\}}{\partial \mu_\ell} \Big|_{\mu} \mu_\ell$$

カーネルの次数 d を明示する場合 K に添え字 d を付けて表す。計算により以下の結果を得ることが出来る。

$$\frac{\partial E_r\{V_\mu(x)\}}{\partial \mu_\ell} \Big|_{o(\mu)} = -K_d(\xi_\ell, \bar{x})$$

次の記法

$$z^{(i)} = (z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n)^t$$

を用い $\bar{\cdot}$ を周辺分布での平均とすれば

$$\frac{\partial K_d(\xi_\ell, \bar{x})}{\partial q_i^k} = k \xi_\ell^i K_{d-1}(\xi_\ell^{(i)}, \bar{x}^{(i)})$$

となり次の結果を得る

$$- \frac{\partial E\{V_\mu(x)\}}{\partial q_i^k} = \sum_{\ell=1}^m k \xi_\ell^i K_{d-1}(\xi_\ell^{(i)}, \bar{x}^{(i)}) \mu_\ell$$

この式から周辺確率を計算することが出来る。

学習には最尤方を用いる事が出来る。学習データ $\{\xi_1, \dots, \xi_m\}$ が与えられたとき次の対数尤度を最大化する

$$L(\mu) = \sum_{\ell=1}^m \log P_\mu(\xi_\ell)$$

最適値では,

$$\frac{\partial L(\mu)}{\partial \mu} = 0$$

平均場を用いて計算すると

$$\frac{\partial L(\mu)}{\partial \mu_{\ell'}} = \sum_{\ell=1}^m (K(\xi_{\ell'}, \xi_{\ell}) - K(\xi_{\ell'}, \bar{x}))$$

我々得られる。以上のアルゴリズムを画像のカテゴリ識別に適用し他の有力な学習アルゴリズムと比較した。学習とテストに用いたデータは Caltech101 であり以下のサイトからとれる。

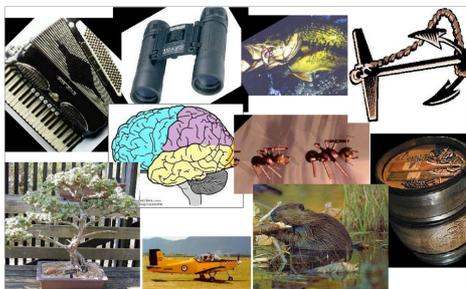
[http://www.vision.caltech.edu/Image_Datasets/Caltech101/

Caltech101.html]

画像の特徴量は SIFT と grid-SIFT を用いた。Bag of Words におけるコードワード数は 200, grid-SIFT の画素間隔を 10pixel とした。以下が識別結果である。

識別器	識別率 1	識別率 2	識別率 3
DHDP	43.3%		
SVM	46.7%	48%	47.3%
KMF	50%	51.3%	

但し、識別率 1-3 は、それぞれ、SVM では線形カーネル、シグモイドカーネル、多項式カーネル、KMF では対称多項式カーネル次数 2, 4 に対応する。以下にサンプル画像と用いた 10 のカテゴリ名を示す。



1	2	3	4	5
accordion	airplane	anchor	ant	barrel

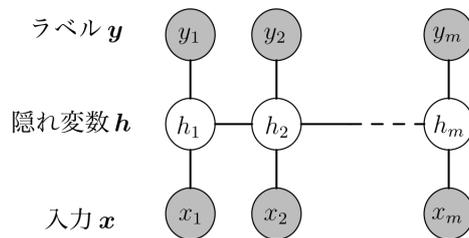
6	7	8	9	10
bass	beaver	binocular	bonsai	brain

次の成果は、平均場近似により直接ランダムフィールドをモデル化し、識別モデルとして用いて、ビデオ画像認識に応用したものである。識別結果は世界最高精度を誇る。

・識別モデル

時系列を扱う代表的な手法は 隠れマルコ

フモデルであるが、提案手法は隠れマルコフモデルより表現力に優れている。また学習データの欠損に対しても、双方向的な結合であるためこれを補う性質がある。提案手法は、以下のように図式化されるモデルである。



ここで

$$P(\mathbf{h}|\mathbf{x}, \theta) = \frac{1}{Z(\mathbf{x}, \theta)} \exp\left(\sum_k \theta_k \cdot \mathbf{F}_k(\mathbf{h}, \mathbf{x})\right)$$

$$Z(\mathbf{x}, \theta) = \sum_{\mathbf{h}} \exp\left(\sum_k \theta_k \cdot \mathbf{F}_k(\mathbf{h}, \mathbf{x})\right)$$

$$F_k(\mathbf{h}, \mathbf{x}) = \begin{cases} \sum_{t=1}^{T-1} t_k(h_t, h_{t+1}) \\ \sum_{t=1}^T \sum_{l=1}^L \sum_{w=-W}^W s_k(h_t, x_{t+w,l}) \end{cases}$$

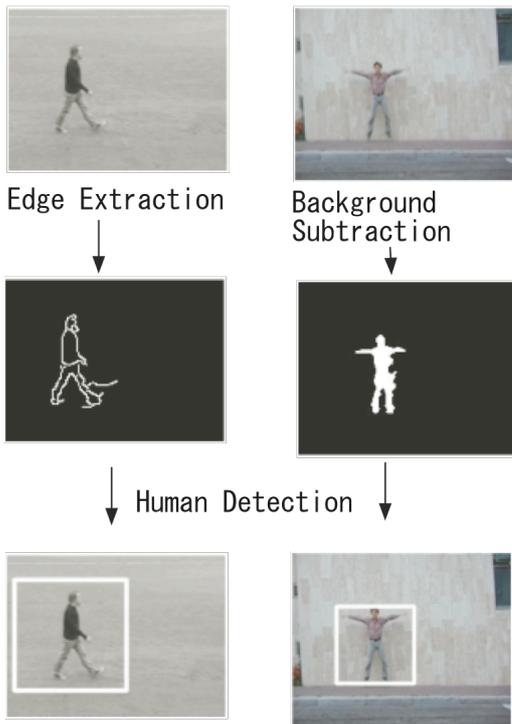
$$s_{\alpha,l}(h_j, \mathbf{x}_i) = \begin{cases} 1 & : \text{if } h_j = \alpha, \quad x_i = l \\ 0 & : \text{otherwise} \end{cases}$$

$$t_{\alpha,\beta}(h_t, h_{t+1}) = \begin{cases} 1 & : \text{if } h_t = \alpha, \quad h_{t+1} = \beta, \\ & \alpha \in H_{y_t}, \beta \in H_{y_{t+1}} \\ 0 & : \text{otherwise} \end{cases}$$

パラメータ推定は最尤推定を用いて行う。具体的には以下の関数を最大とする値を計算する。

$$L(\theta) = \sum_{n=1}^N \log P(\mathbf{y}_n | \mathbf{x}_n, \theta) - \frac{1}{2\sigma^2} \|\theta\|^2$$

学習にはビデオ画像から特徴抽出を行う前処理の過程を含む。前処理では、まず画像中の対象物の検出を行う。検出はエッジ検出を行いその情報から対象物(今の場合人間)を推定し、正方形で囲む処理である。その後はこの正方形の中だけに着目し、運動を表す特徴量をとる。この過程を以下に図示する。また長方形を推定するアルゴリズムも示す。アルゴリズム中、X, Y は正方形の中心を表し、R はその一辺のサイズを表す。



R_0, X_0, Y_0 : initial values
for $t = 1$ **to** T **do**
 $X_t = \text{mean of } X$;
 $S_t = \{(X, Y) : X_t - \frac{1}{4}R_{t-1} \leq X \leq X_t + \frac{1}{4}R_{t-1}\}$
 $X_t = \text{mean of } X \text{ in } S_t$;
 (modified mean by resampling)
 $Y_{max} = \text{maximum value of } Y \text{ in } S_t$;
 $Y_{min} = \text{minimum value of } Y \text{ in } S_t$;
 $R_t = Y_{max} - Y_{min}$;
 $Y_t = \frac{1}{2}(Y_{max} + Y_{min})$;
if $|R_t - R_{t-1}| > 0.15R_{t-1}$ **then**
 R_t is taken to be 15% smaller or larger than R_{t-1} ;
end
end

アルゴリズム

・特徴抽出

特徴量の抽出する前に、動画中の各画像をバイラテラルフィルタを用いて、平滑化した。このフィルタは、画像中のエッジを保持したまま平滑化が行えるのが特徴である。このフィルタでは、近傍領域 W の画素との距離と輝度値の差、それぞれにガウス分布で重みを付け平滑化する。平滑化でノイズ除去をした後は、動画中の人の位置、大きさを求め正方形のウィンドウで切り抜く。人の位置検出は、使ったデータによって使った手法が少し違う。データについての詳細は 4 章で記述するが、Weizmann データセットには背景差分によるマスク画像が付属しており、それを元に人の位置を計算した。KTH データセットでは、背景差分等は使わず直接画像から位置を推定する。こちらのデータでは、幸い背景に強いエッジがないデータであっ

た。そのため、背景から強いエッジを計算することで人の位置を検出することができた。実際には元画像に対してメディアンフィルタを 2 度かけることで、背景の弱いエッジを極力減らした後、キャニー法によりエッジを求めた。ただこれだけだと、背景に水平方向のエッジが残ることがある。そのため、このエッジ画像を水平方向に微分し、なるべく背景のエッジを減らした画像を計算して人の位置推定に使う。KTH データセットの場合は、計算したエッジ画像を用いて人の位置を推定する。Kanade-Lucas-Tomasi 法は、オプティカルフローを求める手法の 1 つである。この手法では追跡に適した特徴点を検出し、その点のオプティカルフローだけを計算する。そのため、比較的ノイズ等に強く高精度なオプティカルフローが得られる。また、特徴点のオプティカルフローだけを扱うため特徴量の大きさを抑えることができる。

提案手法の性能評価は、人の行動認識実験で一般的に使われる KTH データセットと Weizmann データセットを用いて行う。また各実験では、画像の各フレームから最終的に 10 点の特徴点を選び、そのオプティカルフローを 7920 段階に離散化し、特徴量とする。この実験では、提案手法は全て C++ を用いて実装し、CPU:Core2Duo 3Ghz, MainMemory:2Gbyte の計算機を用いて実行した。

実験結果を以下に示す。テストデータと書かれた表では、学習に用いるトレーニングデータとテストデータをあらかじめ分離し実験を行った結果である。またクロスバリデーションと書かれた表は、すべてのデータを学習に用いるが 10 データの内 1 種類だけを抜き出して学習に用いずにテストに用いる。これをすべての種類に対し適用し、最後に平均をとる。

テストデータ

Method	Training Data	Recognition Rate
Our Approach	1/3	0.9348
Schndler[14]	4/5	0.927
Fathi[11]	2/3	0.905
Ahmad[13]	1/3	0.8883

クロスバリデーション

Method	Recognition Rate
Our Approach	0.9569
Gilbert et al[4]	0.967
Kim et al[3]	0.95
Lin et al[12]	0.9343

なお、参考のために、簡単なデータである Weizmann データセットにおける結果を示す。

Weizmann dataset

Method	Recognition Rate
Our Approach	0.957
Fathi[11]	1.00
Lin[12]	1.00
Blank[10]	1.00

以上より、本成果が最も識別率の良い結果を与えていることが分かる。しかも、計算は実時間で出来るため、更に実用的な成果であると言える。

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 9 件)

- ① Lameswar Debnath, Haruhisa Takahashi and Takio Kurita, "A comparison of SVM-based evolutionary methods for multicategory cancer diagnosis using microarray gene expression data," Journal of Systematics, Cybernetics and Informatics, 査読有, pp21-25, (2011).
- ② Rameswar Debnath and Haruhisa Takahashi, "A comparison of SVM-based criteria in evolutionary method for gene selection and classification of microarray data," World Academy of Science Engineering and technology, 査読有, issue 71, pp. 403-407, (2010).
- ③ Hirokazu Nagai, Haruhisa Takahashi, and Kazuhiro Hotta, "Fast Human Action Recognition Using Conditional Random Field," Journal of Signal Processing, 査読有, Vol. 14, No. 6, pp499-505, November (2010).
- ④ Atsuo Nomoto, Kazuhiro Hotta, and Haruhisa Takahashi, "Asbestos Detection in Building Materials Through Consolidation of Similarities in Color and Shape Features," Journal of Robotics and Mechatronics, 査読有, JRM Vol. 22 No. 4 Aug. pp. 496-505, (2010).
- ⑤ J. O. Mercado, K. Hotta, H. Takahashi, M. N. Miyatake, K. T. Medina, H. P. Meana, "Improving the Eigenphase Method for Face Recognition," IEICE Electronics Express, 査読有, Vol. 6, No. 15, pp. 1112-1117, (2009).
- ⑥ Yoshitaka Moriguchi, Kazuhiro Hotta, Haruhisa Takahashi, "An Asbestos Detection method from Microscope Images using Support Vector Random Field of Local Color Features," IEEJ Trans. EIS, 査読有, Vol. 129, No. 5, pp. 818-823, (2009).
- ⑦ 鈴木有紀, 堀田一弘, 高橋治久, "文脈情報に基づく対象が存在する事前確率の推定," 電気学会論文誌, 査読有, Volume 129-C, Number 5, pp. 832-837, MAY (2009).
- ⑧ K. Kawabata, S. Morishita, H. Takemura, K. Hotta, T. Mishima, H. Asama, H. Mizoguchi and H. Takahashi, "Development of an Automated Microscopic System for Asbestos Qualitative Analysis by Dispersion Staining Method," Journal of Robotics and Mechatronics, 査読有, Vol. 21 No. 2, pp. 186-192, Apr. (2009).
- ⑨ R. Debnath, M. Muramatsu and H. Takahashi, "Implementation issues of second-order cone programming approaches for support vector machine learning problems" IEICE Trans. Fundamentals, 査読有, vol. E92-A, no. 43. pp. 1209-1222, (2009).

[学会発表] (計 16 件)

- ① Manabu Yoshida, Haruhisa Takahashi, "Human Action Recognition with Optical Flow and Space-Time Patch", RISP International Workshop on Nonlinear Circuits, 2, Honolulu, Hawaii, March 4-6, (2012).
- ② 杉山正暁, 高橋治久, "クラスタリングとグラフラプリアンを用いた画像分類", DEIM2012, 2012, 3月 02-05, シーサイドホテル舞子ビラ神戸.
- ③ Jesus Robles-Castro, Gonzalo Duchon-Sanchez, Haruhisa Takahashi: "Improving object position estimation based on non-linear mapping using Relevance Vector Machine." CONIELECOMP 2011: Feb 28, 2011 - Mar 2, 2011, San Andres Cholula, Mexico, 171-17.
- ④ H. Takahashi, "Kernel Random field and Mean field approximation", The third Int. Symp. On robot and Artificial Intelligence, Shanghai Jiao Tong University, (2010)
- ⑤ Rameswar Debnath and Haruhisa takahashi, "A comparison of SVM-based criteria in evolutionary method for gene selection and classification of microarray data," World Academy of Science Engineering and technology, issue 71, pp. 403-407, ID 64, (2010).
- ⑥ Rameswar Debnath and Haruhisa takahashi, "A new ensemble learning with support vector machine," In proceedings of the 2010 International Conference on Computer and Information Application, Tianjin, China, December, pp 33-35, (2010).
- ⑦ Kouichi Aso, Haruhisa Takahashi and

Kazuhiro Hotta, ” Recurrent Temporal Restricted Boltzmann Machine for Motion Recognition”, 2010 International Workshop on Nonlinear Circuits, Communication and Signal Processing NCSP'10, Waikiki, Hawaii, March 3-5, pp 512-515, (2010).

⑧ Shinnosuke Nomoto, Haruhisa Takahashi and Kazuhiro Hotta, ” The Text Regions Extraction from General Scenes Using Edge Comparing and Support Vector Machine”, 2010 International Workshop on Nonlinear Circuits, Communication and Signal Processing NCSP'10, Waikiki, Hawaii, March 3-5, pp 612-615, (2010).

⑨ Hirokazu Nagai, Haruhisa Takahashi and Kazuhiro Hotta, ” Fast Human Action Recognition Using Conditional Random Field”, 2010 International Workshop on Nonlinear Circuits, Communication and Signal Processing NCSP'10, Waikiki, Hawaii, March 3-5, pp 648-651, (2010).

⑩ Ryohei SEKIGUCHI, Haruhisa TAKAHASHI and Kazuhiro HOTTA, ” The Automatic Parameter Tuning for Multi-class Learning with KDA “ 2010 International Workshop on Nonlinear Circuits, Communication and Signal Processing NCSP'10, Waikiki, Hawaii, March 3-5, pp 190-193, (2010).

⑪ T. Tanaka, K. Hotta and H. Takahashi, ” Object Categorization Based on Probabilistic Integration of Local and Global Features, ” 7th IASTED International Conference on Signal Processing, Pattern Recognition and Applications (SSPRA2010), Innsbruck, Austria, Feb. 17-19 pp. 325-332, (2010).

⑫ Moriguchi, K. Hotta and H. Takahashi, ” Accuracy Improvement of Asbestos Detection by Rectangular Neighboring Region for SVRF of Local Color and Shape Features, ” 7th IASTED International Conference on Signal Processing, Pattern Recognition and Applications (SSPRA2010), Innsbruck, Austria, Feb. 17-19, pp. 221-227 (2010).

⑬ T. Ishihara, K. Hotta and H. Takahashi, ” Estimation of Object Position Based on Color and Shape Contextual Information”, Proc. of 15th International Conference on Image Analysis and Processing (ICIAP2009), Vol. 5716, pp. 825-834, Salerno, Italy, May, (2009).

⑭ K. Kawabata, H. Yamazaki, Y. Tsubota, T. Mishima, K. Hotta, H. Asama, H. Mizoguchi and H. Takahashi, ” Development of an Automatic Polarized Microscopic Imaging System for Asbestos Qualitative

Analysis, ” IEEE/ASME Conference on Advanced Intelligent Mechatronics (AIM2009), Singapore, Jul. 14-17 (2009).

⑮ Y. Tsubota, K. Kawabata, H. Yamazaki, T. Mishima, K. Hotta, H. Asama, H. Mizoguchi and H. Takahashi, ” Development of a Polarized Microscopic Image Management System for Supporting Asbestos Qualitative Analysis utilizing Dispersion Staining, ” 35th Annual Conference of the IEEE International Conference of Industrial Electronics Society (IECON2009), Porto, Portugal, Nov. 3-5 2009.

⑯ Y. Mimura, K. Hotta and H. Takahashi, ” Action Recognition Based on Estimation of Conditional Probability by Non-parametric Model, ” International Symposium on Visual Computing (ISVC2009), Lecture Notes in Computer Science, Vol. 5867, Las Vegas, USA, Nov. 30-Dec. 2, pp. 489-498 (2009).

〔図書〕 (計 1 件)

(1) 高橋治久 堀田一弘 ” 学習理論”, コロナ社 (215 ページ) (2009)

6. 研究組織

(1) 研究代表者

高橋 治久 (TAKAHASHI HARUHISA)
電気通信大学・大学院情報理工学研究科・
教授
研究者番号：90135418

(2) 研究分担者

堀田 一弘 (KAZUHIRO HOTTA)
名城大学・理工学部・准教授
研究者番号：40345426
庄野 逸 (HAYARU SHOUNO)
電気通信大学・大学院情報理工学研究科・
准教授
研究者番号：50263231

(3) 研究協力者

Debnath Rameswer (DEBNATH RAMESWER)
Khulne University (Bangladesh) ・
Computer Sc. & Engineering Discipline ・
Professor