

平成30年6月13日現在

機関番号：13904

研究種目：挑戦的萌芽研究

研究期間：2015～2017

課題番号：15K12027

研究課題名（和文）マルチモーダルな観察データに基づく動植物鑑定システム

研究課題名（英文）Flora and fauna identification system based on multi-modal data observed

研究代表者

青野 雅樹（Aono, Masaki）

豊橋技術科学大学・工学（系）研究科（研究院）・教授

研究者番号：00372540

交付決定額（研究期間全体）：（直接経費） 2,700,000円

研究成果の概要（和文）：インターネット上に投稿された植物写真画像から、その学名を鑑定することを目的とした国際コンテストPlantCLEFに研究期間2年目から参加し、2016年に世界第一位、2017年も、「雑音を含む訓練データ」トラックで世界第一位、全体で世界第二位を獲得した。技術的には、撮影場所や時間などのメタデータ特徴量に加え、Fisher vector等の伝統的な特徴量と、深層学習の畳み込み層や多層パーセプトロン層の特徴量を混合することで精度向上に成功した。また、端数型プーリング法を開発し、正方形画像だけでなく任意の矩形画像での畳み込みを可能とした。

研究成果の概要（英文）：From the plant photo images posted on the Internet, We have participated in the international contest called PlantCLEF from the second year of the research period, resulting in No.1 world record in 2016 and the year 2017, "Noisy training data" truck, which was the second place as a whole. Technically, we have succeeded in improving the accuracy by mixing textual features from meta data with the image features of CNN and MLP layers of the deep learning as well as the traditional features such as Fisher vector. Furthermore, we have developed a fractional pooling method to enable convolution with arbitrary rectangular images.

研究分野：マルチメディア データマイニング

キーワード：植物鑑定 深層学習 画像特徴量

1. 研究開始当初の背景

高度コンテンツ創造・分析・流通技術は我が国の UNS 戦略プログラムで重点研究開発課題に指定されているという、大まかな背景があった。一方、我々を取り囲む環境として、画像検索・自動アノテーション付与等を目的とした国際コンペ ImageCLEF (www.imageclef.org) で 2014 年から、「動植物の膨大なデータ(画像・動画・音声・テキスト)を学習し、未知な動植物を鑑定するタスク」(通称 LifeCLEF) が始まり、国際的に、『動植物の正確な鑑定タスク』への関心が高まっており、生物の多様性の理解と持続可能な人間社会の構築を全世界の人が共有できる環境を作ろう、という世界的なトレンドがあった。

人間よりもはるかに小さい動植物には類似する品種が多く、植物の「葉」に代表される容易に撮影できる対象からの鑑定は植物学者でさえも困難で挑戦的課題であった。研究代表者は、ImageCLEF 国際コンテストの「画像への自動アノテーション」タスクに 2013 年から参加し、2 年目の 2014 年に世界第一位の精度を達成することができた。このタスクの参加者には、日本からは我々の他に東大、九大、国立情報学研究所 (NII) チームが参加、海外からは、フランス、スペイン、ロシア等、全体で 40 以上研究機関が参加し、最終的に課題を提出できた 12 チームで競い、世界最高精度を達成することができた。この背景のものと、画像のアノテーションで養った技術を動植物の鑑定技術に活かしたいと意欲を燃やしていた。

2. 研究の目的

上記の背景の下、マルチモーダルな非侵襲性 (DNA 鑑定のように侵襲しない) データから「動植物の鑑定タスク」において世界最高レベルの技術を開発することを研究目的とした。本研究の斬新性とチャレンジ性は、一般にサイズのばらつきが大きく、類似種類間の差異が小さい動植物に関して、「カラー写真、ビデオ、テキスト、地理上の位置を含むメタ情報」などマルチモーダルなデータから、「これは何?」という質問に正確に答えられる技術の開発と設定した。

3. 研究の方法

マルチモーダルなアノテーション技術をもとに、細粒度かつ高精度な鑑定手法を開発し、最終的に、デモシステムを公開した。まず、高精度な鑑定技術の研究開発に向けて、様々な手法を開発した。

その技術をもとに、国際コンテストである、PlantCLEF に 2016 年と 2017 年の 2 年連続で参加した。以下、その研究方法を詳述する。

まず、PlantCLEF データは 2016 年時点で、約 11 万 3000 画像(1000 種類の植物)あり、内訳は「枝」、「葉」、「花」、「果実」、「樹皮」のように、植物だけでも多様な画像データの混合であり、これに加えて「人間の足」「全く植物と関係ない画像」などのノイズが多数含まれる。2016 年時点で植物の種類が 1000 種類なので、単純な分類問題としては、ImageNet と同様で 1000 種類のクラスからなる。しかし、ImageNet と異なり、基本的に、非常に似通った植物画像が多数組あること、また、ノイズ画像が多数含まれる点も、この国際コンテストの課題を挑戦的なものにしていく。

我々が取り組んだのは、ノイズ画像を事前に識別するための仕組みである (学会発表論文[15])。2017 年になり、10 倍の 10,000 種類の植物画像となり、50 層の ResNet をもとに、端数型の MaxPooling 法を開発し (学会発表論文[8])、さらに分類画像の解像度を徐々にダウンサンプリングする新手法で分類の高精度化を達成した (雑誌論文[1])。

4. 研究成果

結果として、2016 年ならびに 2017 年の植物の学名の鑑定を競う国際コンテストに 2 年連続で世界一の鑑定精度を達成できた。

(1) PlantCLEF2016 国際コンテストで鑑定精度世界最高性能 (学会発表論文[15])

ここでは、基本的には ImageNet で訓練済みの 16 層の VGG-16 を利用した。ただし、本来の VGG-16 は入力画像として 224×224 という小ぶりの画像に固定する必要がある。これに対して PlantCLEF2016 で提供されたデータは任意サイズの画像で正方形の画像はほとんどない。実際に画像のアスペクト比は図 1 のように分布している。これは植物画像に限らず、この種のデータ処理では必須と考えら

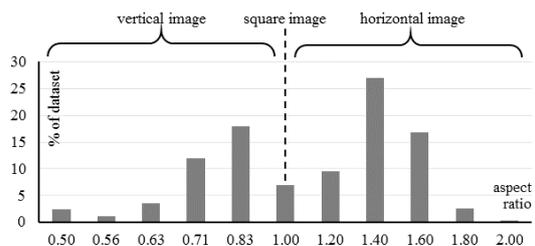


図 1 植物画像のアスペクト比

具体的な方法として、我々は VGG-16

の最後の Pooling 層 (Conv5_3 層) を Spatial Pyramid Pooling (SPP) とした。これにより、畳込みの feature map のサイズは、1×1, 2×2, 4×4 の合計 21 領域に拡張でき、個々の領域を Average Pooling すると、元来の 512 種類の feature map が $21 \times 512 = 10752$ 種類の固定長のベクトルとなり、任意サイズの画像の入力に対して、最終層の Pooling でサイズの違いを吸収できるようにした。次に、活性化関数としては、負のデータをゼロ値化する ReLU が過学習を抑制するために一般的だが、ReLU より柔軟で学習時にダイナミックに負の部分の係数を学習できる PReLU (Parametric ReLU) を採用した。

通常の ImageNet のような国際コンテストと PlantCLEF の大きな違いは、分類すべきクラスが通常の植物である 1000 種類だけでなく、“Unseen Category” (どこにも分類されないカテゴリー) を用意しないといけない点である。

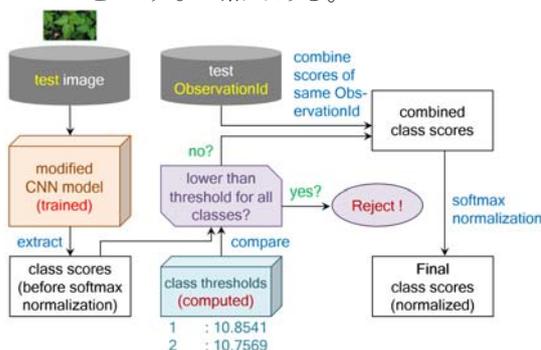


図2 提案システム (訓練時)

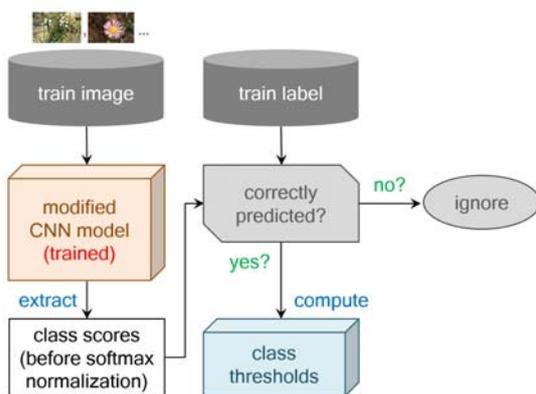


図3 提案システム (テスト時)

図2と図3が提案したシステムである。まず、通常の VGG-16 のような畳込みを何層かでやる CNN モデルを用意したあと、上述のように、最終層を SPP 式のプーリング方式としたものを modified CNN model と書いている。訓練時は、答えであるラベルと比較し、正確に予測できていたクラスだけを残し、「クラスごとの閾値」

計算を行う。テスト時は、このクラスごとの閾値を用いて、“Unseen Category” の正確な予測を行う。同時にテスト時には、観測された ID (Observation ID) を利用している。これは、同時刻に観察された植物は、高い確率で同種類の植物であるとの理由で植物クラスの推定に利用している。

上述のアルゴリズムを主として適用し、2016 年の PlantCLEF2016 で世界最高性能を達成した。

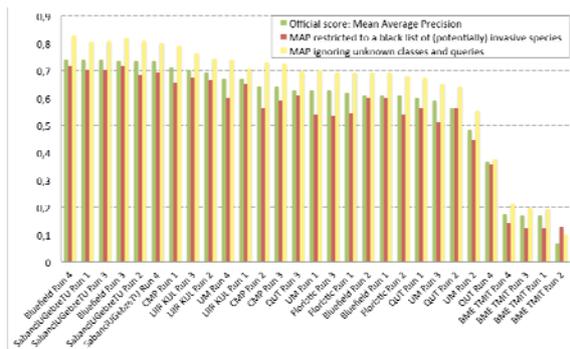


図4 PlantCLEF2016 の性能結果。左端の第一位が我々のチーム。我々のチーム名は Blue Field である。
(<http://www.imageclef.org/lifeclef/2016/plant> 参照)



図5 人間でも間違いやすいが、提案システムで正しく植物の学名を識別できた成功例

(2) PlantCLEF2017 でも雑音環境データで世界最高性能を達成 (学会発表論文[8])

2017 年度の国際コンテストでは、植物の種類が 1000 から 10000 と 10 倍に増加し、同時に人間の足とか、洋服のパターンとか、植物とはおよそ無関係な画像が雑音データとして含まれた。訓練データ側も、信頼できる訓練データと雑音のある訓練データに分けられた。この両方を使ったデータではチームとしては世界第 2 位であったが、雑音のある訓練データだけを使った結果では、世界第 1 位の成果だった。

PlantCLEF2017 で我々は、50 層ある Residual Network を主として利用し、これに、種々の改良を施した深層学習手法を開発した。

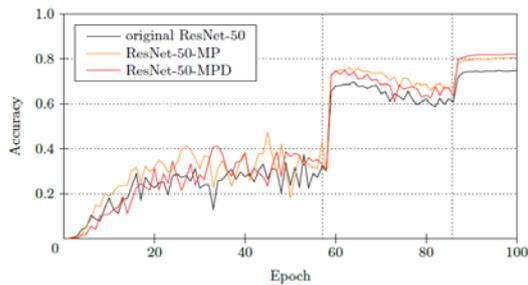


図6 Epoch-Accuracy グラフ (赤と橙色が提案法)

特に、Max Pooling に関して我々は端数型の新手法を開発した。この端数型 Max Pooling に加えて畳込みにおけるダウンサンプリングを遅延させることで最終的な分類精度を向上させることに成功した。これを Max Pooling Delayed (MPD) 手法と命名し、50 層の ResNet に組み込んで実験した結果が図6である。エポックが少ないときには、ばらつきはあるものの、何もしない従来法よりも分類精度が向上していることがわかる。

(3) 端数型の Max Pooling 手法の開発 (雑誌論文 [1])

通常の Max Pooling では、サイズが $1/2$, $1/4$ のように feature map が急激に小さくなることに伴い、繊細な特徴量が失われてしまう問題が観察される。そこで、端数型の Max Pooling 法を導入し、畳込み層を重ねてダウンサンプリングする際、急激なサイズダウンを避ける工夫をした。たとえば、 224×224 のサイズでスタートしたとしても、以下の図7と図8のように徐々にサイズを小さくすることが可能である。

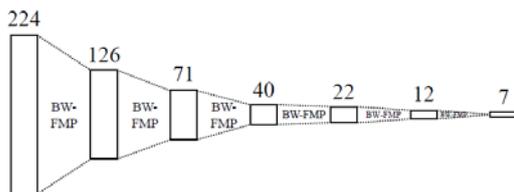


図7 端数型の Max Pooling で5層ではなく6層の畳込みと Pooling により、精度落ちを緩和する例

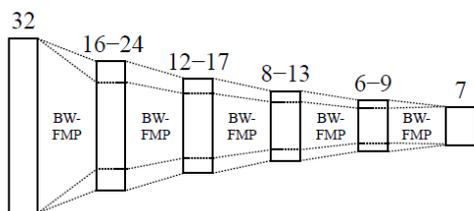


図8 端数型 Max Pooling を 32×32 の画像に適用する事例

BW-FMP とは、ここで提案した手法であり、Bilinear Weighting - Fractional Max Pooling 手法の略称である。PlantCLEF2017 でも、この手法の有用性を確認できたが、植物画像に限らず、FCVC-Aircraft データ (航空機画像の識別データ) や Oxford-IIIT Pet Dataset (猫と犬の画像データセット) などの動物画像の分類性能でも、細粒度の識別性能を発揮することを示すことが出来た。以下の図9と図10に端数型 Max Pooling の仕組みと動物画像への適用例を示す。

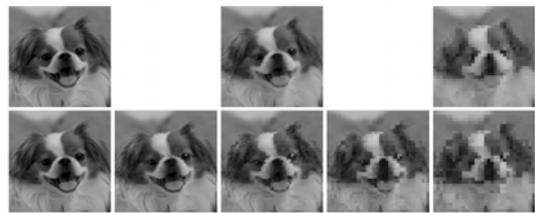


図9 上段は 128×128 の画像に通常の Max Pooling を適用し、 $64 \times 64 \rightarrow 32 \times 32$ とした事例。下段は 128×128 の画像に提案手法を適用し、 $128 \times 128 \rightarrow 90 \times 90 \rightarrow 64 \times 64 \rightarrow 45 \times 45 \rightarrow 32 \times 32$ のように端数型の Max Pooling を適用したもの。通常で得られない解像度の feature map が得られるだけでなく、精度の急降下を緩和することが出来、動植物の鑑定精度を高めることに成功した。

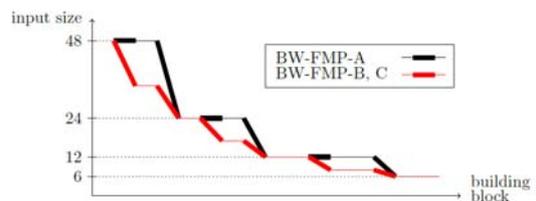


図10 端数型 Max Pooling 法を適用し、解像度の変化を表現したもの (3 種類を例にとったもの)。このうち、太い線が端数型 Max Pooling によるもの。黒い手法 (BW-FMP-A) は通常の Max Pooling を特殊例として含むことを強調したもの。赤色のグラフは解像度の変化が滑らかになるように2種類の乱数で徐々に解像度を低下させたもの。

5. 主な発表論文等

〔雑誌論文〕 (計 5 件)

[1] Siang Thye Hang, Masaki Aono, Bi-linearly weighted fractional max pooling, Multimedia Tools and Applications, Vol. 76, pp. 22095-22117, Springer, 2017, DOI: <http://dx.doi.org/10.1007/s11042-017-4840-5>
 [2] 濱田和真, 立間淳司, 青野雅樹,

Volumetric Spiral Scan Order による三次元考古遺物モデルの類似検索, 情報処理学会論文誌, Vol.10, No.2, pp.1-7, June 2017, ISSN:1882-7799, NAID:170000148699

- [3] Atsushi Tatsuma, Masaki Aono, Food Image Recognition using Covariance of Convolutional Layer Feature Maps, IEICE Transactions on Information and Systems, Vol. E99-D, No. 6, pp. 1711-1715, 2016, DOI:http://dx.doi.org/10.1587/transinf.2015EDL8212
- [4] S. Biasotti, A. Cerri, Masaki Aono, A. Ben Hamza, V. Garro, A. Giachetti, D. Giorgi, A. Godil, C. Li, Chika Sanada, M. Spagnuolo, Atsushi Tatsuma, S. Velasco-Forero, Retrieval and classification methods for textured 3D models: A comparative study, Visual Computer, Vol. 32, Issue 2, pp.217-241, Springer, 2016, DOI:http://dx.doi.org/10.1007/s00371-015-1146-3
- [5] Md Zia Ullah, Masaki Aono, A Bipartite Graph-based Ranking Approach to Query Subtopics Diversification Focused on Word Embedding Features, IEICE Transactions on Information and Systems, pp. 3090-3099, Vol. E99-D, No. 12, 2016, DOI:http://dx.doi.org/10.1587/transinf.2016EDP719

[学会発表] (計 21 件)

- [1] 竹ノ内達哉, 青野雅樹, エッジと輝度反転に着目した特徴量と CNN 特徴量を用いたマンガ画像の著者推定, 電子情報通信学会総合大会, D-6-10, 1p, 東京電機大学, 3月20日, 2018.
- [2] 濱田和真, 青野雅樹, ボクセル群の奥行き密度反映した画像による屋内 3D シーン分類, パターン認識・メディア理解研究会 (PRMU2017-204 (2018-3)), 6pp, 東京都渋谷区 青山学院大学青山キャンパス, 3月19日, 2018.
- [3] 岩淵渉, 青野雅樹, 局所特徴量を用いた 3D CNN による 3 次元モデルの部分検索, 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM2018), I2-1, 6pp, 福井県あわら市清風荘, 3月4日, 2018.
- [4] 杉山裕哉, 青野雅樹, 画像へのマルチタグとそれらの重要度の自動付与, 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM2018), I1-1, 6pp, 福井県あわら市清風荘, 3月4日, 2018.
- [5] Ryo Miyagi and Masaki Aono, Sliced Voxel Representations with LSTM and CNN for 3D Shape Recognition, Asia Pacific Signal and Information Processing Association (APSIPA2017), 4pp, December 13th, Kuala Lumpur, Malaysia, 2017.
- [6] 高垣幸秀, 青野雅樹, Deep Spatio-Temporal Transformation を用いた逐次人物動作検出, パターン認識・メディア理解研究会 (PRMU 2017-68), pp.31-35, 熊本大学, 10月12日, 2017.
- [7] 岩室伸哉, 青野雅樹, 異方性を考慮した画像の符号化に基づく時系列データ分類, パターン認識・メディア理解研究会 (PRMU 2017-56), pp.137-142, 東京大学, 9月16日, 2017.
- [8] Siang Thye Hang and Masaki Aono, Residual Network with Delayed Max Pooling for Very Large Scale Plant Identification, LifeCLEF 2017 Workshop in Conference and Labs of the Evaluation Forum (CLEF2017), 7pp, September 11-14, Dublin, Ireland, 2017.
- [9] Ko Endo, Masaki Aono, Eric Nichols, Kotarou Funakoshi, An Attention-based Regression Model for Grounding Textual Phrases in Images, International Joint Conference on Artificial Intelligence (IJCAI2017), 7pp, August 25, Melbourne, Australia, 2017.
- [10] 宮城諒, 青野雅樹, LSTM と CNN を用いたボクセル表現に基づく三次元形状類似検索手法の提案, パターン認識・メディア理解 (PRMU) 研究会, PRMU 2016-230, pp.203-208, 名城大学, 3月21日, 2017.
- [11] 大塚 達也, 青野雅樹, 単語分散表現と文法的な表現に着目したカスタマーレビューの観点ごとの評価値推定, DEIM2017 (第9回データ工学と情報マネジメントに関するフォーラム), F8-2, 8pp, 高山グリーンホテル, 3月8日, 2017, 学生プレゼンテーション賞受賞.
- [12] Siang Thye Hang and Masaki Aono, Open World Plant Image Identification Based on Convolutional Neural Network, Asia Pacific Signal and Information Processing Association (APSIPA2016), 4pp, December 16th, Jeju, Korea, 2016.
- [13] Md. Zia Ullah, Md. Shajalal, Abu Nowshed Chy, and Masaki Aono, Query Subtopic Mining Exploiting Word Embedding for Search Result Diversification, 12th Asia Information Retrieval Societies Conference (AIRS2016), pp 308-314, November 30th, Beijing, China, 2016 Best Short Paper Presentation Awarded.
- [14] 遠藤 昂, 船越孝太郎, エリック・ニコルズ, 青野雅樹, テキストクエリで指定した画像中の領域を特定する一手法, 電子情報通信学会 画像工学研究会 信学技報, IE2016-66, pp.7-12, 福岡大学, 10月6日,

- 2016.
- [15] Siang Thye Hang, Atsushi Tatsuma, and Masaki Aono, Bluefield (KDE TUT) at LifeCLEF 2016 Plant Identification Task, LifeCLEF 2016 Workshop in Conference and Labs of the Evaluation Forum (CLEF2016), 10pp, September 6th, University of Évora, Portugal, **2016 World No.1 Plant Identification Record in PlantCLEF2016 International Contest**
- [16] M. Savva, F. Yu, Hao Su, M. Aono, B. Chen, D. Cohen-Or, W. Deng, Hang Su, S. Bai, X. Bai, N. Fish, J. Han, E. Kalogerakis, E. G. Learned-Miller, Y. Li, M. Liao, S. Maji, A. Tatsuma, Y. Wang, N. Zhang, Z. Zhou, Large-Scale 3D Shape Retrieval from ShapeNet Core55, 9th Eurographics Workshop on 3D Object Retrieval, May 8, Lisbon, Portugal, 2016.
- [17] 吉井和輝, エリック・ニコルズ, 船越孝太郎, 中野幹生, 青野雅樹, 顕著性マップを用いた画像の説明文自動生成, JSAI 2016, 人工知能学会全国大会, 4K1-5, 4pp, 北九州国際会議センター, 6月9日, 2016.
- [18] 吉井和輝, Eric Nichols, 船越孝太郎, 中野幹生, 青野雅樹, 画像内領域の顕著性を考慮した画像の説明文生成, PRMU 2015-164, 電子情報通信学会パターン認識・メディア理解研究会 信学技法, pp.1-6, 産業総合研究所(臨海副都心センター), 3月24日, 2016.
- [19] 高垣幸秀, 青野雅樹, 時間軸を考慮した特徴量の提案とそれを用いた動画分類, D-12-34, 電子情報通信学会総合大会, 3月16日, 九州大学伊都キャンパス, 2016.
- [20] Shoki Tashiro and Masaki Aono, 3D Shape Retrieval from a Photo Using Intrinsic Image, Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA2015), 4pp, December 19, Hong Kong, IEEE Signal Processing Society, 2015.
- [21] 阿部卓也, 立間淳司, 青野雅樹, 料理レシピサイトから抽出される特徴に基づいた調理時間予測, D-018, 第14回情報科学技術フォーラム (FIT2015), 愛媛大学, 2pp, 9月16日, 2015.

〔図書〕(計 0件)

〔産業財産権〕

○出願状況(計 5件)

- [1] 名称: 三次元画像分類装置および三次元画像分類方法

- 発明者: 青野雅樹, 濱田和真
権利者: 同上
種類: 特許権
番号: 特願 2018- 046791
出願年月日: 平成 30 年 3 月 14 日
国内外の別: 国内
- [2] 名称: 検出装置および検出方法
発明者: 青野雅樹, 高垣幸秀
権利者: 同上
種類: 特許権
番号: 特願 2017- 197694
出願年月日: 平成 29 年 10 月 11 日
国内外の別: 国内
- [3] 名称: 画像特徴量及びそれを用いる三次元形状検索システム
発明者: 青野雅樹, 立間淳司, 鈴木将也
権利者: 同上
種類: 特許権
番号: 特願 2017-042156
出願年月日: 平成 29 年 3 月 6 日
国内外の別: 国内
- [4] 名称: 画像認識装置、画像認識方法、及び画像認識プログラム
発明者: 青野雅樹, 立間淳司
権利者: 同上
種類: 特許権
番号: 特願 2016-008273
出願年月日: 平成 28 年 1 月 19 日
国内外の別: 国内
- [5] 名称: 画像特徴量及びそれを用いる三次元形状検索システム
発明者: 青野雅樹, 立間淳司, 鈴木将也
権利者: 同上
種類: 特許権
番号: 特願 2017-042156
出願年月日: 平成 29 年 3 月 6 日
国内外の別: 国内

○取得状況(計 1件)

名称: データのインデックスの次元削減方法及びそれを利用したデータ検索方法並びに装置
発明者: 青野雅樹, 立間淳司
権利者: 同上
種類: 特許権
番号: 特許第 5818023 号
取得年月日: 平成 27 年 10 月 9 日
国内外の別: 国内

〔その他〕
ホームページ等

6. 研究組織

(1) 研究代表者

青野 雅樹 (AONO, Masaki)

豊橋技術科学大学・大学院工学研究科・教授

研究者番号: 00372540