科学研究費助成事業　　研究成果報告書

平成　２６　年　　６　月　２０　日現在

機関番号：　３２７０８
研究種目：　基盤研究(C)
研究期間：　2011～2013
課題番号：　２３５００２３７

研究課題名（和文）シーン内の文脈情報を利用した高速画像分類手法の実現

研究課題名（英文）Fast Image Categorization Method using Scene-Context Scale Information

研究代表者
姜　有宜（Kang, Yousun）

東京工芸大学・工学部・准教授

研究者番号：１０５８２８９３

交付決定額（研究期間全体）：（直接経費）　　3,800,000円、（間接経費）　　1,140,000円

研究成果の概要（和文）：本研究で提案したシーン内の新しい文脈情報シーンコンテキストスケール（Scene-Context Scale）を用いて実際の画像データベース（MSRC21とPASCAL）に適用し、画像分類及び画像のセグメンテーションの実験を行った。高速特徴抽出器であるランダムフォレスト（Random Forest）を使うことでより高速に評価し従来手法に比べ、大幅な分類精度の向上が得られることを確認した。さらに、開発したフレームワークをロボットの物体認識の研究にも応用し、空間コンテキスト情報を基づいて物体認識を行った。すべての研究成果は取りまとめ、国内と海外の学会にその成果を発表した。

研究成果の概要（英文）：We propose scale-optimized textons to learn the best scale for each object in a scene. We incorporate them into image categorization and semantic segmentation. Our textonization module produces a scale-optimized codebook of visual words. We approach the scale-optimization problem of textons using the scene-context scale in each image, which is the effective scale of local context to classify an image pixel in a scene. We perform the textonization process using a randomized decision forest, which is a powerful tool with high computational efficiency in vision applications. Results of our experiments using MSRC21 and VOC 2007 segmentation datasets demonstrate that our scale-optimized textons improve image categorization and segmentation performance.

研究分野：　情報学

科研費の分科・細目：　知覚情報処理・知能ロボティクス

キーワード：　画像分類　画像セグメンテーション　シーンコンテキストスケール

１．研究開始当初の背景

(1) 最近、デジタルカメラの普及に伴ってWeb アルバムのように膨大なデータベースから画像を探索する作業が多くなってきた。そのため、より早く、より正確及び簡単に画像を分類する研究が注目されている。個人ユーザにより多くのデジタルイメージが共有されている今の時代に、自動的に画像を分類できるシステムの開発は、最も重要な研究テーマである。

(2) 多くのコンピュータビジョン研究において、コンテクスト（context）情報が認識性能を高めることが示されている。しかし、コンテクストと言う概念を明確に定義することは容易なことではない。なぜなら、人間の脳がどのように物体を認識するのかがまだ明確に知られていないからである。今でも生理学の分野で多くの研究者によって人間の認識システムに影響を与える新しいコンテクスト源の報告がなされている。このような中で、我々は様々なコンテクスト情報の中で、シーンコンテクストスケール（Scene-Context Scale）と言う新たなコンテクスト情報に着目し、画像分類に適用した。シーンコンテクスト（Scene-Context）情報は、物体認識とシーン解析の研究において重要な手がかりを与える。

２．研究の目的

(1) 画像分類には大きく二つの種類がある。一つ目は画像全体を認識し、その一枚の画像がどんなカテゴリーに属しているかを判断するシーン分類（scene categorization）である。これに使われるカテゴリーには、山、海、都市、部屋、教室などがある。二つ目は画像中の個々の物体を認識してから画像を分類するものである。この分類では、犬、猫、馬、牛、車、人など様々な物体のカテゴリーが使える。本研究では画像の中に存在する物体を認識してから画像分類を行う後者の研究を目的にする。

(2) これまで様々なシーンコンテクスト情報が開発され研究に適用してきたが、本研究ではシーンコンテクストスケール（Scene-Context Scale）という新しいコンテクスト情報を開発し、それを用いて自動的に画像分類を行う。従来、画像の特徴記述を行う際に、物体のスケールに関係なく画像の局所的な領域から特徴を抽出していた。このため、物体のスケール変化に対して頑強な特徴表現を得ることが困難であった。そこで我々はシーンコンテクストスケールという概念を導入し、画像毎のシーンコンテクストスケールを予め求めることで、物体のスケールに応じたシーンコンテクスト情報を利用する画像分類（Image Categorization）手法を開発するのを研究の目的にする。

３．研究の方法

(1) マルチスケールテクストンフォレスト（Multi-scale Texton Forest；MTF）：画像から抽出した特徴量の密度が高いほど高性能な画像分類を実現することができる。テクストン（texton）は代表的な高密度の特徴量であり、近年ではテクスチャ解析や一般物体認識にも特徴量として多く用いられ、その有効性が確認されている。画像から特徴を抽出し、その特徴からテクストンを生成する段階の作業はテクストナイジェーション（textonization）と呼ばれている。最近ランダムフォレストを用いてテクストナイジェーションを行うセマンティックテクストンフォレスト（Semantic Texton Forest；STF）が提案されている。本研究では、STF のスケールを拡張して得られるマルチスケールテクストンフォレスト（Multi-scale Texton Forest；MTF）を生成してシーンコンテクストスケールを推定する。MTF を生成する具体的な方法は、まず、図 1 の左に示すように、入力画像から d×d のサイズのイメージパッチ p を取り出し、取り出した関心領域内に存在するピクセルをランダムに選び、決定木のノード分岐にそのピクセル値を用いる。図 1 の右にある関数 f (v)は足し算、引き算、ピクセルの絶対値など単純計算を行うことで、多くの決定木のノードが高速に分岐できる。h はランダムに選ばれた閾値である。

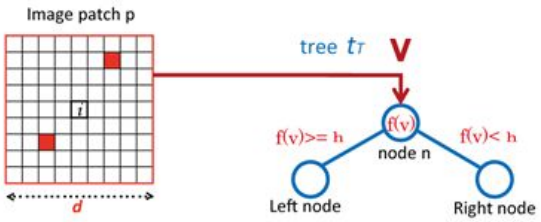

図１　イメージパッチ p と決定木のノード

我々は MTF を生成するため、イメージパッチ p のサイズを増加し、スケールレベルを拡張する方法を使う。ピクセルごとに有効なローカルコンテキストの範囲を決定するため、様々なスケールでテクストナイジェーションを行い、一番効果的なスケールを求める。MTF は様々なスケールを持つ STF の集合であり、全体スケールスペースを S と定義すると各スケールステップは であり、i は 1 から k まで定義される。

(2) シーンコンテクストスケール（Scene-Context Scale）：決定木のノードを分岐することでセマンティックなテクストンを探し、葉ノード (leaf node)に到達するとイメージパッチ p は階層的にクラスタリングされる。多数のサンプリングによって葉ノードではクラス分布の計算ができる。その分布の平均値を求めることでカテゴリーの識別が可能になる。MTF は様々なスケールにおいてテクストンを生成し、各スケールごとにカテゴリー分布も得られる。その分布をピク

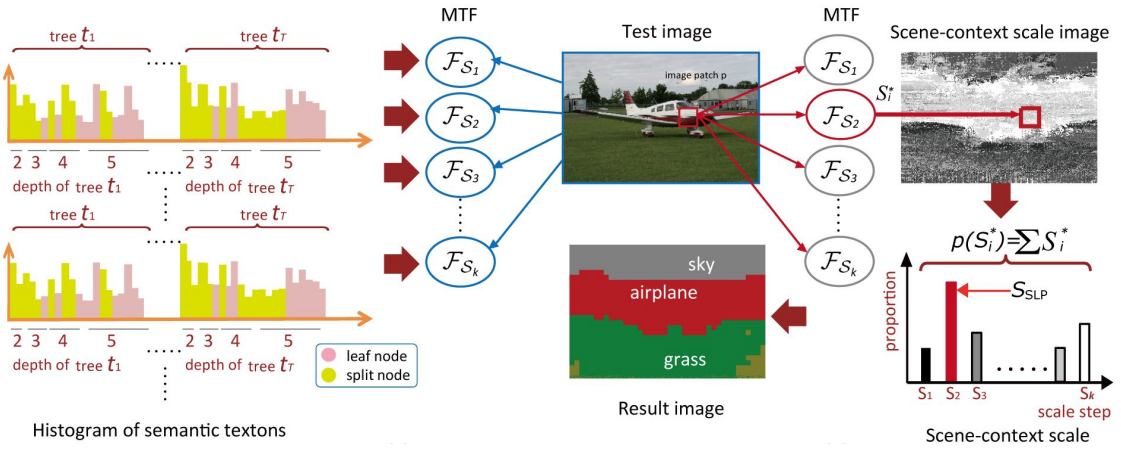

図２　シーンコンテクストスケールを用いたカテゴリーの識別

セルごとに求め、一番有効なスケールステップを探し、それをシーンコンテクストスケールとする。MTF から得られたマルチスケールテクストンは様々なスケールを持っている画像特徴である。その特徴を用いてシーンコンテクストスケールを推定する。効果的なローカルコンテキストの範囲を決めるシーンコンテクストスケールは、画像のピクセルごとに求める。その理由は、物体が画像ごとに違うスケールを持っているからである。有効なローカルコンテキストの範囲をシーンコンテクストスケールと呼び、全スケールスペース上で一つのスケールステップを求める。MTF を構成している個々の STF 葉ノードから各カテゴリ分布を求めた後、そのエントロピーを計算する。エントロピーは個々の葉ノードから計算されることから、MTF のエントロピーもスケール毎に求められる。計算されたエントロピーの中、最も低いスケールステップを一つ選び、そのスケールスペースをシーンコンテクストスケールと決める。そして、画像全体のシーンコンテクストスケールの分布は、取り出したすべてのイメージパッチのシーンコンテクストスケールを求めることで分かる。

(3) Pyramid　Match Kernel：提案手法では、画像分類の手段として bag of　keyword と呼ばれる手法を用いる。ここでの keyword とは、多様なスケールを持っているテクストンの特徴である。図 2 に示すように、個々の STF から作られたヒストグラムは、STF を構成している決定木の分岐ノード(split node)と葉ノードそれぞれに対する度数で構成される。階層的なクラスタ表現という観点から、葉ノードのみの度数だけではなく、分岐ノードの度数分布も併せて表現する。これらの度数分布を用いて画像分類を行うため、本稿では非線形サポートベクトルマシン(SVM)を使う。非線形 SVM は、カーネルの選択によって汎化性能が大きく異なる。そこで、ヒストグラムの類似度に基づく Pyramid　Match Kernel を用いた。また、１対他の２クラス分類を組み合わせることで多クラス識別システムを構成した。はじめに、個々の STF に対して非線形 SVM による分類を行い、各カテゴリー分布を求める。続いて、シーンコンテクストスケールに基づいて各カテゴリー毎の分布を再計算する。

４．研究成果

(1) 実験環境

画像分類の実験に用いたデータセットは、マイクロソフト研究所で開発された MSRC セグメンテーションデータセットである。このデータセットには 21 種類のオブジェクトが 591 枚の画像に亘ってピクセル単位でラベル付けされている。21 種類のオブジェクトは、建、草、木、牛、羊、空、飛行機、水、顔、車、自転車、花、標識、鳥、本、椅子、道路、猫、犬、人のボディ、ボートである。実験では、多様な背景や照明、スケールなどを持つ画像から、学習用として 265 枚の画像、評価用として 250 枚の画像を使用した。まず、MTF(Multi-scale Texton Forest)を学習するため、スケールの異なる６つの STF(Semantic Texton Forest)を用意する。すなわち、異なる６種類のサイズのイメージパッチを学習画像から切り出すことで、６つの STF を別々に学習する。イメージパッチの初期サイズを 15×15 とし、スケールステップが 6 段階なのでイメージパッチのサイズは 90×90 まで増やす。一つの STF には 5 本の木が含まれ、1 個の木の最大深さは 10 である。切り出したイメージパッチのサンプル数は、 の場合 500 枚で、スケールが増加するとサンプル数も増しておく。処理時間は STF が 500 枚のサンプル数を持つ場合、その STF を学習するのに約 30 秒かかり、テストには一枚の画像に対して 0.1 秒を要した。

(2) 画像クラスタリングにおける成果

MTF を用いたクラスタリング結果を図 3 に示す。異なる 6 つの STF を使用したため、その結果も 6 列で示されている。テストした元画像と正解画像は、1 列目と 2 列目に示されており、一番小さいスケールで作成された決定木によって得られたクラスタリング結果は 3

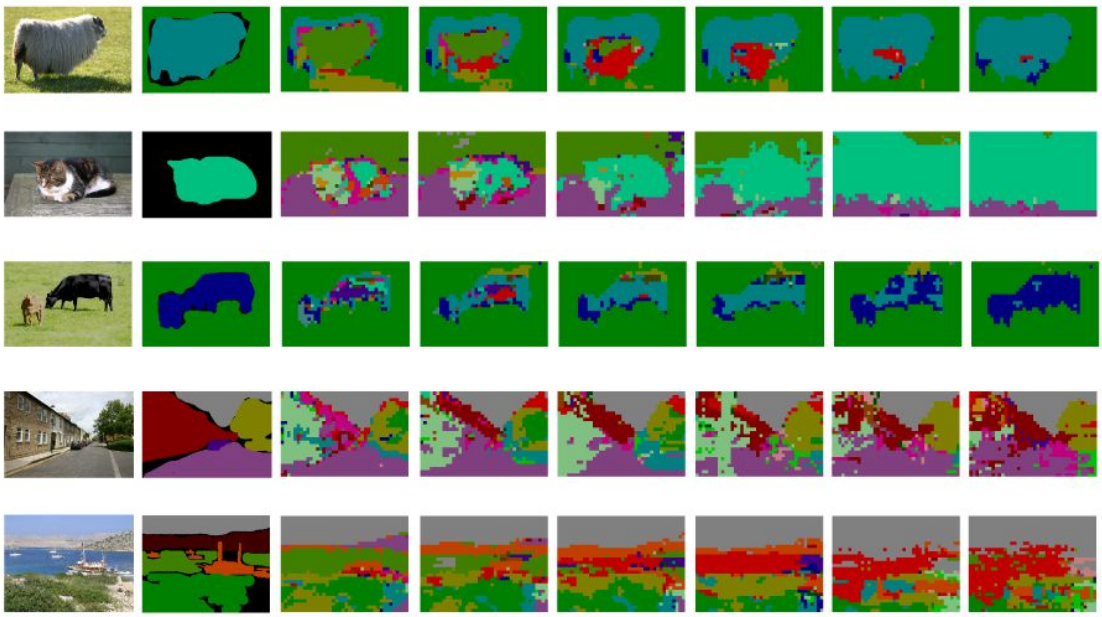

図 3　マルチスケールテクストンフォレストを用いたクラスタリングの結果

列目に表わされている。右の列に行く程、そのスケールステップは大きくなる。クラスタリングされた個々のピクセルは、各決定木の最後のノードに到達した時点で、カテゴリー毎に確率が与えられる。その確率の平均を求めることで、最もありそうなカテゴリが推定される。図 3 のように、それぞれの結果を色で表せば、局所的なカテゴリの分類と同時に画像のセグメンテーションも可能である。
図 3 のクラスタリング結果を見ると、1 行から 3 行までの画像については、より大きいスケールで学習した方が良い結果を示している。しかし、4 行目と 5 行目の画像は、予想されるように背景自体がオブジェクトになり、最も小さいスケールからクラスタリングしたほうがよるよい結果を示している。これは入力画像によって適切なシーンコンテキストスケールを求め、その確率を画像分類やセグメンテーションに適用する必要があることを意味している。そして、従来手法と比較してみるとシーンコンテキストスケールを用いてクラスタリングを行ったほうが、クラス平均精度もグローバル精度もよい性能が得られている。特に、グローバル精度は従来研究の 38.4%から 48.3%に向上した。
(3) 画像分類における成果
シーンコンテキストスケールを用いて画像分類を行った結果は表 1 に示す。表 1(a)はシーンコンテキストスケールを用いていなかった従来研究の結果であり、(b)は SLP(Scale Level Prior)を用いて計算した結果である。(c)はただの分布平均値から求められた分類結果である。そして表 1(d)シーンコンテキストスケールを用いた分類結果である。我々の提案方法によって、(a)の全カテゴリ平均値 72.8%より 74.9%まで認識性能が向上した。特に提案方法では、3 つのクラスを除いたすべてカテゴリに対して認識性能が改善された。これらの結果より、本論文で提案したシーンコンテキストスケールは、画像のカテゴリー分類においても強力及び効果的なコンテキスト情報であることが明らかとなった。
(4) 本研究では、シーンコンテキストスケール（Scene-Context Scale）と呼ばれる新しいシーン内の文脈情報（Scene Context）を提案し、その文脈情報に基づいて大規模の画像データベースから画像分類及び画像検索を実現することであった。研究開発 1 年目には、ランダムフォレスト（Random Forests）を用いて、画像ごとにスケール最適化されたテクストンを求めた。ランダムフォレストをマルチスケール（Multi-scale）に拡張し、そこから画像ごとに正しいシーンコンテキストスケールを推定することでスケール最適化されたテクストンの抽出が高速にできた。2 年目には、求めた画像の特徴量からローカル特徴量とグローバル特徴量に分け、その特徴量を上手く統合するために新しいマルチカーネル学習（Multiple kernel learning）手法を提案した。異なるスケール空間から抽出された画像のローカル特徴量とグローバル特徴量をより効果的に統合するため、新しいマルチカーネルのモデルを構築した。提案されたマルチカーネル学習手法は、画像分類及び画像のセグメンテーションに適用し、実験でその評価を行った。3 年目には、提案したシーン内の新しい文脈情報と学習方法を実際の画像データベースに適用し、画像分類の実験を行いその評価を発表した。画像データベースはチャレンジングな大規模のデータベース、The PASCAL Visual Object Classes Challenge 2007 を利用した。さらに、開発したフレームワークをロボットの物体認識研究技術に適用し、空間コンテキスト情報に基づいて物体認識を行った。3 年間開発した技術は、取りまとめ国内と海外の会議や著名な学術誌に投稿し、成果の発表を行った。

	building	grass	tree	cow	sheep	sky	aeroplane	water	face	car	bicycle	flower	sign	bird	book	chair	road	cat	dog	body	boat	class average
(a) None	64	86	75	**86**	**92**	90	74	66	64	**88**	72	84	70	**53**	90	**67**	67	57	**36**	64	**77**	72.8
(b) SLP	63	93	**81**	82	73	**97**	64	**75**	**84**	61	**73**	**85**	**72**	49	93	56	**77**	57	35	**84**	64	72.2
(c) Mean	**71**	88	80	83	77	95	**87**	70	73	86	71	83	67	**53**	**94**	62	71	59	33	74	70	73.7
(d) Distribution	64	**94**	79	84	82	**97**	83	**75**	**84**	79	**73**	**85**	59	49	91	58	75	**68**	32	**84**	**77**	**74.9**

表1　シーンコンテキストスケールによる画像分類の結果

５．主な発表論文等

〔雑誌論文〕（計　6　件）

①Yousun Kang and Akihiro Sugimoto, Image Categorization and Semantic Segmentation using Scale-Optimized Textons, IT Convergence Practice, 査読有, Vol.2, No.1, 2014, pp. 2-14, http://isyou.info/inpra/papers/inpra-v2n1-01.pdf

② Duk Shin, Hiroyuki Kambara, Natsue Yoshimura, Yousun Kang, and Yasuharu Koike, Control of a Brick-breaking Game using Electromyogram, International Journal of Engineering and Technology, 査読有, Vol.6, No.2, 2014, pp.128-131, http://www.ijetch.org/papers/680-W00043.pdf

③ Tam T. Le, Yousun Kang and Akihiro Sugimoto, Image Categorization Using Hierarchical Spatial Matching Kernel, The Journal of the Institute of Image Electronics Engineers of Japan, 査読有, Vol.42, No.2, 2013, pp.214-221, http://www.iieej.org/mokuji/mokuji-42-2.pdf

④姜有宣, 長橋　宏, 杉本晃宏, 画像分類のためのランダムフォレストを用いた シーンコンテキストスケールの開発, 画像ラボ, 査読無, Vol.23, No.8, 2012, pp.5-11, 日本工業出版, http://www.nikko-pb.co.jp/nk_comm/mok08/html/images/1208g05.pdf

⑤ Yousun Kang, Hiroshi Nagahashi, and Akihiro Sugimoto, Image Categorization Using Scene-Context Scale Based on Random Forests , IEICE Trans. on Information and Systems, 査読有, Vol.E94-D, No.9, 2011, pp.1809-1816 http://search.ieice.org/bin/pdf.php?lang=E&year=2011&fname=e94-d_9_1809&abst=

⑥ Yousun Kang, Koichiro Yamaguchi, Takashi Naito, and Yoshiki Ninomiya, Multiband Image Segmentation and Object Recognition for Understanding Road Scenes, IEEE Trans. on Intelligent Transportation Systems, 査読有, Vol.12, No.4, 2011, pp.1423-1433, DOI: 10.1109/TITS.2011.2160539

〔学会発表〕（計　10　件）

① Yousun Kang, Layout Estimation and Object Recognition Using Spatial Context Information, Workshop on General Intelligence for Humanoid Robots in ICRA2014, 2014年6月1日, Hongkong

② Yousun Kang, Illumination Invariant Face Detection Using Multiband Camera System, International Conference on Smart Media Applications (SMA), 2013年10月15日, Kota kinabalu, Malaysia

③ Yousun Kang, Multiband Camera System using Color and Near Infrared Images, International Conference on Electronics, Mechatronics and Automation (ICEMA), 2013年8月25日, Singapore, Win the Prize of the Best Presentation.

④Duk Shin, Control of a Brick-breaking Game using Electromyogram, The 3th International Workshop on Computer Science and Engineering, 2013年8月25日, Singapore

⑤ Yousun Kang, Face Detection using Multiband Camera System, IEEE International Conference on Multimedia and Expo (ICME), 2013年7月18日, San Fransisco, CA

⑥姜 有宣, マルチバンドカメラを用いた顔検出システム, 情報処理学会 電子化知的財産・社会基盤研究会 (EIP), 2013年5月16日, 横浜

⑦Yousun Kang, Texton Clustering for Local Classification using Scene-Context Scale, Japan-Korea Joint Workshop on Frontiers of Computer Vision (FCV), 2013年1月31日, Incheon, Korea

⑧ Yousun Kang, Scale-Optimized Textons for Image Categorization and Segmentation , IEEE International Symposium on Multimedia (ISM), 2011年12月6日, Dana Point, CA

⑨Yousun Kang, Road Image Segmentation and Recognition using Hierarchical Bag-of-Textons Method , 5th Pacific Rim Symposium (PSIVT), 2011年11月22日, Gwangju, Korea

⑩Tam T. Le, Hierarchical Spatial Matching Kernel for Image Categorization, International Conference on Image Analysis and Recognition (ICIAR), 2011年6月23日, Vancouver, Canada

〔図書〕（計　0　件）

〔産業財産権〕
○出願状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

６．研究組織
(1)研究代表者
姜　有宣（KANG, Yousun）
東京工芸大学・工学部・准教授
研究者番号：10582893