様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２８　年　　６　月　　４　日現在

機関番号：　１７１０２
研究種目：　基盤研究(B)（一般）
研究期間：　2012～2015
課題番号：　２４３０００７４

研究課題名（和文）実社会の「今」を切り取るプロキシ型情報収集機構に関する研究

研究課題名（英文）A proxy mechanism to capture a present situation of the society

研究代表者
谷口　倫一郎（Taniguchi, Rin-ichiro）

九州大学・システム情報科学研究科（研究院・教授

研究者番号：２０１３６５５０

交付決定額（研究期間全体）：（直接経費）　　13,800,000円

研究成果の概要（和文）：本研究では，社会において，情報を必要とする者（情報需要者）と情報を提供する者（情報供給者）をつなぐ情報伝達の新たな枠組みの確立を目指した．特に，モダリティを変換することで第三者への情報伝達を効率化することを狙って，画像アノテーション技術を中心に研究を進めた．具体的には，画像アノテーションの性能改善のための相関トピックモデルの構築，不完全なトレーニングデータからの画像アノテーションモデル構築，異なるモダリティのデータを相互検索できるCross-modal Retrievalの構築，撮影プロセスのモデルを基にした複数画像の要約等を行った．これらの成果は関連分野の主要国際会議，国際誌等で発表した．

研究成果の概要（英文）：In this research, a new communication mechanism between people who require some information and people who can provide the information. We have focused on how to utilize image media which can be easily acquired by smartphones, We suppose transformation of modality can be applied to achieve an efficient and effective information transfer. Especially, image annotation techniques can be used as modality transformation mechanism, and we have researched mainly to develop new algorithms related to image annotation, We have constructed correlated topic model to improve image annotation performance, image annotation model from incomplete training data set, cross-modal retrieval framework realizing mutual retrieval of multi-modal data, and summarization of users' captured images based on a model of human's image capturing process. These research results have presented in several top-level international conferences and journals.

研究分野：　コンピュータビジョン

キーワード：　クロスモーダル検索　画像アノテーション　画像検索

１．研究開始当初の背景

テキストや画像など様々な情報を大量に蓄積する仕組みが構築されてきており，インターネットを通じてそれらの情報に容易にアクセスして様々な情報を容易に入手できるようになってきた．昨今では，ソーシャルネットワーキングサービス（SNS）の普及により，社会で起こった出来事や話題のニュースなど最近の情報も容易に入手できるようになってきている．しかし，そこで得られる情報はあくまでもデータが記録された時点，すなわち「過去」に基づく情報であり，「今」に関する情報を得ることはできない．「今」のデータを取得するには，従来のようにデータ提供者の都合でデータをインターネット上等に蓄積していく枠組みとは異なる新たなデータ収集の枠組みが必要である．

一方で，携帯電話（特にスマートフォン）の普及により，世界のいたるところで，簡便に実世界の情報を獲得し，インターネット上に提供する基盤が整ってきており，潜在的には「今」に関する情報を獲得できるようになってきていると考えられる．そこで本研究では，多数のスマートフォンを実世界の「今」を切り取るセンサとして有効に活用し，ユーザの欲する情報を獲得する新たな情報獲得の枠組みについて研究を行う．そこでは，情報を欲する側（情報需要者）と情報を提供する側（情報供給者）の需給関係をスムーズに整合させる仕組みが必要不可欠であり，本研究ではその点に焦点をあてて研究を行う．

情報供給者から実際に投稿してもらう情報には，スマートフォンで撮影した「今の写真」を想定している．テキストベースの情報を提供してもらうと，情報供給者が「今」を見てそれをテキストに変換するという部分に情報供給者の主観が含まれてしまう恐れがある．情報収集の公平性という観点からすれば，情報供給者の主観情報は極力排除されるべきであり，そのためにスマートフォンで「今」という候補を「撮影する」という形式を採用する．従って，情報供給者から提供される多くの情報から必要な情報を抽出・統合する処理機構が求められる．

これまでの情報収集の仕組みと本研究で目指す情報収集の仕組みを情報需要者の観点から整理すると次のようになる．「過去」に基づく情報収集では，情報供給者が投稿した情報がデータベースに蓄積されており，情報需要者は自分の欲する情報がデータベースに蓄積されていることを期待してキーワードをベースとした情報検索をする．つまり，情報需要者からみれば所望の情報は自動的にデータベースに収集されてくることを期待した，いわば「Passive 型（受動型）」であると言える．これに対して，「今」に関する情報を得るには，情報需要者自身が，自分の欲する情報が収集されるように情報収集のトリガを引くという「Active 型（能動型）」である必要がある．実際に情報を収集するのは情報提供者であるため，情報需要者は情報供給者に対して情報収集の意図を的確に伝える必要がある．また，情報供給者から投稿された情報の中から，自らが真に欲する情報を整理・統合する必要がある．情報需要者から情報供給者への情報収集の意図を伝達することから始まり，情報供給者から提供された情報から情報需要者の必要な情報を抽出するという作業すべてを，情報需要者に課すのは容易ではないため，この情報伝達の仕組みを効率よく行う仕組みを確立する必要がある．

２．研究の目的

まず，第一に情報需要者の意図を情報供給者へ伝達する仕組みについて研究を進める．従来の Web 検索のように単純なキーワードベースで情報需要者の意図を情報供給者に伝達する方法では，次のような問題が生じる可能性がある．

・キーワードが多すぎると情報供給者が意図を理解するのが難しくなる．
・キーワードから連想される情報が情報需要者と情報供給者間で一致しているとは限らない．

そこで本研究では，キーワードに変わる手段として，画像を効果的に利用して意図を伝達する仕組みを明らかにする．また，情報需要者の情報収集意図には，プライバシーに係わる情報が潜在している可能性も考慮すべきであるため，プライバシー保護の観点から意図を適切に隠蔽して伝達する仕組みについても明らかにする．

第二に，情報供給者から投稿された情報から情報需要者の意図に適合した情報を抽出する仕組みについて研究を進める．基本的に，情報供給者から提供される情報には次のような性質がある．

・情報供給者が投稿する情報は，投稿場所，投稿時間が一致しているとは限らない．
・情報需要者の意図する情報が必ずしも含まれているとは限らない．

そこで，複数の情報供給者から寄せられた情報を整理し，その中から情報需要者の欲する情報を抽出する方法を明らかにする．

３．研究の方法

２．で述べた研究の目的を達成するために以下の研究を進めた．

(1) 画像を効果的に利用して意図を伝達する仕組みについて，特に，画像アノテーション（画像の内容を表すラベルを与えること）の観点から画像から意図につながる重要なキーワードの検出について研究を進めた．画像アノテーションに関する研究は多く行われているが，その性能はまだ十分な

ものとはいえず，その性能向上が不可欠である．本研究では性能向上を目指して，画像アノテーションのための相関トピックモデルの構築について研究を行った．

(2) 画像アノテーションのモデル構築には大量のトレーニングデータ（画像とそれに含まれる対象のラベル群の集合）を用いた学習過程が必要不可欠である．しかし，一般には，サンプルデータは完全なものではなく，誤りが含まれている．そこで，不完全な学習データを用いても正しく画像アノテーションモデルが構築できるようなの構築について研究を進めた．

(3) 将来的には，情報需要者と情報供給者間のコミュニケーションは様々なモダリティ（様式）を用いることが想定される．そこで，画像アノテーション技術をマルチモーダルなデータに対しても適用できるような仕組み（Cross-modal Retrieval），特に，画像からテキストを，あるいはテキストから画像を連想する仕組みについて研究を進めた．

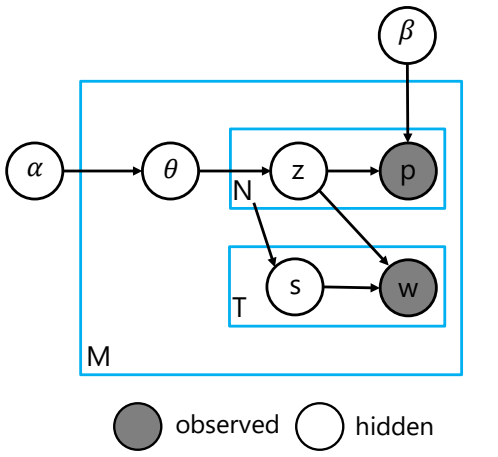
(4) 情報供給者から寄せられる情報の統合については，「画像群の要約」という問題に定式化して研究を進めた．すなわち，画像に記録された位置情報や画像特徴の一貫性などから特に重要度の高い画像を選抜する方法を明らかにした．本手法は，要約する際にどのような観点で要約するかを制御することが可能になっているため，用途に応じた要約が可能になる．

４．研究成果

(1) 画像アノテーションの性能改善

画像の内容は，トピックと呼ばれる仮想的な要素から構成され，トピックは画像特徴あるいはテキスト情報から構成されるというトピックモデルを想定する（図１）．従来は，トピック間の相関を無視していたが，一般には，ある意味のある画像ではトピックに関連があると考えるのが自然である．そこで，本研究では，トピック間に相関を仮定した「相関トピックモデル（Correlated Topic Model）」を構築した（図２）．実験では，図２に示すような単純な相関トピックモデル(CorrCTM)以外にも，いくつかの相関トピックモデルを実現し，相関を想定しないトピックモデルよりも性能が高いことを実験で明らかにした．モデルパラメータは，トレーニングデータを用いて学習によって推定する．

実験では，LabelMe，PASCAL VOC07，Corel 5K という３つの標準的なデータセットを用いて，F1 という標準的な評価基準に基づいて性能を比較した．その結果をグラフで示したものが図３であり，明らかに相関トピックモデルを導入した方（末尾が CTM になっているもの）の性能が向上していることが分かる．図４に実際のアノテーションの例を示しているが，この例でも，明らかに相関トピックモデルの方が適切なキーワードを提示していることが分かる．



- α hyper-parameter of topic
- θ topic distribution for one image
- β prob. of feature point for topic
- z topic assigned to feature point
- s topic assigned to text word

図１　画像のトピックモデル

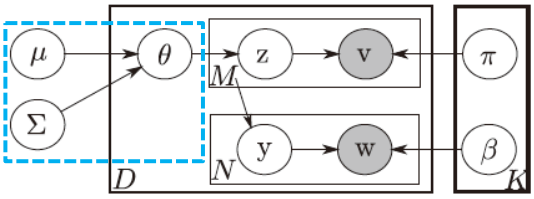


図２　画像の相関トピックモデル

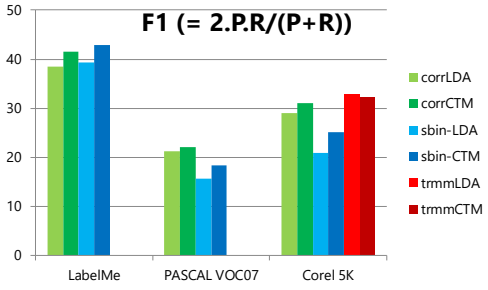


図３　相関トピックモデルの性能評価

Ground truth: sky, sea, sand, people

corrLDA: sky, sea, water, road, car
sLDA-bin: sky, mountain, water, ground, stone
corrCTM: sky, sea, beach, sand, people
sCTM-bin: sky, people, sand, sea, mountain

図４　相関トピックモデルによる画像アノテーション

(2) 不完全データを用いた画像アノテーションモデルの構築

画像アノテーションを実現するためには，事前にトレーニングデータを用いて，アノテ

ーションモデル内のパラメータを推定する必要がある．様々な画像に対してアノテーションを実現するためには，大量のトレーニングデータ（画像とそこに含まれるキーワード群の集合）が必要になる．従来は，トレーニングデータは完全なものであることを前提としていたが，大量データを集めた場合には現実的ではない．そこで，本研究では，トレーニングデータに付与されているラベルが完全ではない（ラベルが欠けている，誤ったラベルが与えられている）状況での，アノテーションモデル構築に関する研究を行った．

具体的には，トレーニングデータの画像とラベル（tag）の関係が図５左のように，画像によっては不完全であるものから，図５右のように，抜け落ちたキーワードとの関係を修復することがゴールである．

本研究ではトレーニング画像 x_n がラベル y_c に関連する程度 R を３つの観点，視覚的な関連性，ラベルの関連性，画像とラベルの関連性を用いて評価する．

- 視覚的な関連性 R_V は，x_n がとその最近傍 $x^* \in \mathcal{X}_c^+$（$\mathcal{X}_c^+$ はラベル c を持つトレーニング画像の集合）との距離で評価する．
- ラベルの関連性 R_S は，x_n に与えられたラベルの集合 Y_n の中で最も共起性が高いラベルとの共起度で評価する．
- 画像とラベルの関連性 R_s は，x_n を近傍として持つ $\mathcal{X}_c^+$ の画像数で評価する．

具体的には下記のように定義している．

$$R(x^n, y) = average(R_V(x^n, y) + R_S(x^n, y) + R_N(x^n, y))$$

$$R_V(x^n, y_c) = 1 - dist(x^n, x^*)$$

Visual similarity based

$$R_S(x^n, y_c) = \max_{y \in Y^n} co\_occur(y_c, y)$$

Semantic similarity based

$$R_N(x^n, y_c) = \sum_{m=1}^{M} \frac{p_m}{m} / \sum_{m=1}^{M} p_m + \varepsilon$$

Reversed nearest neighbor based

- $\mathcal{X} = \{(x^1, Y^1), ..., (x^N, Y^N)\}$: the training set.
- $\mathcal{Y} = \{y_1, ..., y_C\}$: the label set.
- $\mathcal{X}_c^+$ contains images with label y_c, $\mathcal{X}_c^-$ otherwise

この処理によりラベルの割当がどのように変更されるかの例を示したのが図６である．評価はベンチマーク用のデータセットである EPS Game，IAPRTC 12，MIR Flickr を用いて行った．トレーニング用データからラベルを削除する割合を変更した場合に，アノテーションの性能がどのように変化するかを検証した結果を図７に示す．図７の横軸はラベルを削除した割合であり，提案手法（proposed）が従来手法よりも若干ではあるが優れていることが分かる．

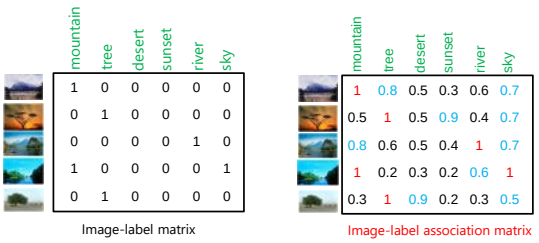
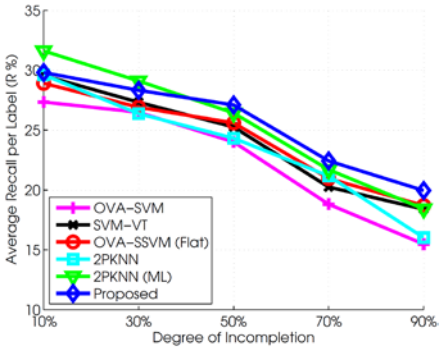


図５ 画像－ラベルの関連性修復の概念

Image x^n	Incomplete labels Y^n	Predicted labels Y	
		OVA-SSVM (Flat)	Proposed
	bloom, leave	bloom, flower, fruit, forest, branch	bloom (0), leave (0), trunk (0.531), flower (0.552), plant (0.765)

図６ 不完全なラベルの修正例



(b) Comparison of R

図７ 不完全トレーニングデータを用いた画像アノテーション性能

(3) クロスモーダルなデータ検索

様々な情報を効果的に取り扱うためには，異なるモダリティ間（Cross Modality）で情報を検索する技術が必要不可欠である．そこで本研究では，画像アノテーションの技術を Cross Modal な情報検索，具体的には画像とテキスト情報を対象とした検索に拡張した．ここでの問題は，画像とテキストでは情報の質が異なる，つまり画像は連続的で密な情報であるのに対し，テキストは疎な情報である点である（図８）．

本研究では，トレーニングデータに基づく学習を２段階で行う．まず，トレーニングデータとして与えられた画像とそれに対応するテキストの特徴情報が同じ潜在空間に写像されるように学習を行う（図９：Coupled Dictionary Learning）．その上で，トレーニングデータに与えられたカテゴリー情報（キーワード）に基づいて分類できるように特徴情報をキーワード空間に写像する（図９：Coupled Feature Mapping）．

情報の検索は，学習フェーズで得られたパラメータを用いて，検索クエリーをキーワード空間に写像し，写像された特徴を用いてデータベースを検索して，特徴とマッチングするデータを出力する（図 10）．このフレームワークは画像からテキストの検索，テキストから画像の検索，いずれも同じである．

提案手法の性能評価は，Pascal Voc，Wiki，MIR Flickr 25K の 3 つのベンチマークデータを用いて行った．評価指標として平均精度（Mean Average Precision，MAP）等を用いた結果を図 11 に示すが，提案手法（proposed）が従来手法よりも優れていることが分かる．また，典型的な検索例を図 12 に示す．

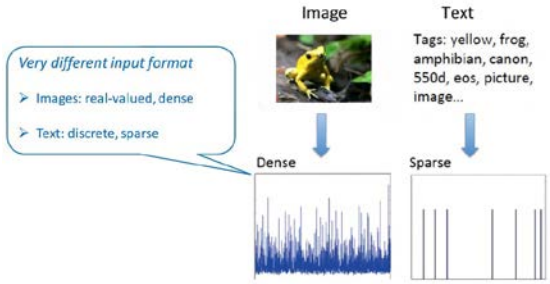


図 8　異なるモダリティでの特徴分布の違い

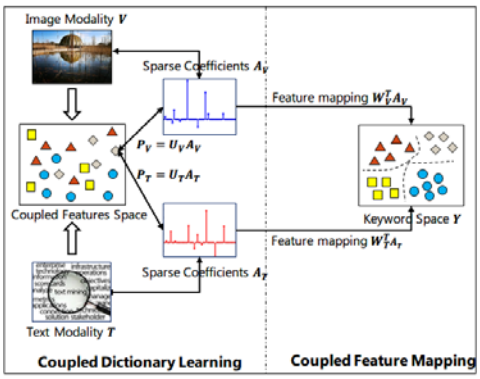


図 9　複数モダリティデータでの学習過程

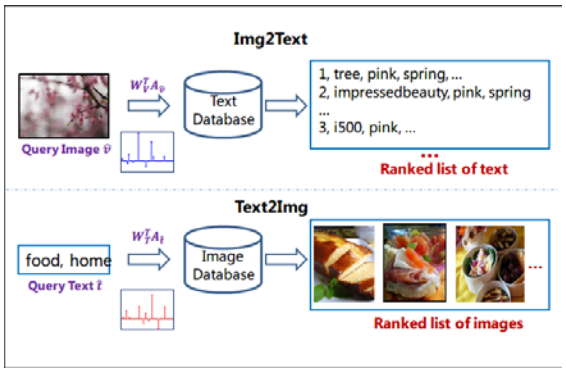


図 10　複数モダリティでの情報検索

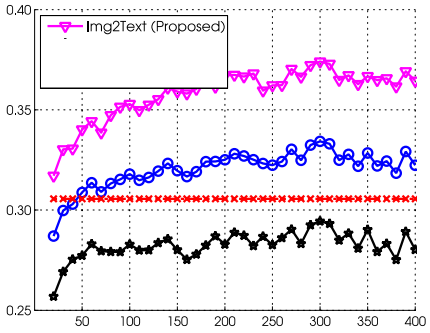


図 11　複数モダリティ検索の性能評価

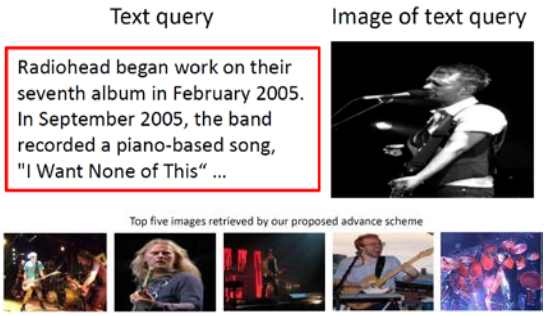


図 12　Text2Image の例

(4) 画像群の要約

スマートフォンの普及に伴い，手軽に多くの画像が撮影され，それがインターネットなど共有スペースに蓄積されるようになってきた．蓄積された多量の画像データを有効に利活用するためには，類似の画像をひとまとめにするといった「要約」の技術が必要である．本研究では，スマートフォンで撮影された画像の要約について，以下の 2 つの仮定に基づいた手法を開発した．

・重要な対象はその撮影回数が多い．
・重要な対象周辺では滞留時間が長い．

基本的な手法は画像の Visual Rank の計算に基づいている．Visual Rank は Web ページの Page Rank の考え方を画像集合に応用したものであり，画像間の類似度に基づいて計算される．値の大きな画像ほど重要な画像であるということになっている．本研究では下記のように Visual Rank を計算する．

$$\boldsymbol{R} = \alpha \boldsymbol{S}\boldsymbol{R} + (1-\alpha)\boldsymbol{I}$$

ここで $\boldsymbol{S}$ は画像間の類似度を表す類似度行列，$\boldsymbol{I}$ は上述の仮定を反映させるバイアス項である．$\boldsymbol{I}$ は，次のように定めた．

$$\boldsymbol{I} = \gamma \boldsymbol{T} + (1-\gamma)\boldsymbol{P}$$

T は滞留時間を表す項であり，***P*** は画像の撮影枚数を表す項である．

上記で求めた Visual Rank では，画像の類似性が支配的になる傾向があるので，ランク r の画像とランク(r+1)の画像の類似性が高い時には ***I*** 値が低い方を棄却するといった re-ranking 処理を付加している．

詳細な実験結果は省略するが，おおよそ，本手法で得られた要約画像の半数は人手でも選ばれているが，半数は異なっている．特に，滞留時間が長いところで撮った画像が必ずしも要約の対象にならない場合があることが分かる．より意味的な解釈に踏み込んだ処理が必要と思われる．

５．主な発表論文等

〔雑誌論文〕（計　3 件）

① Xing Xu, Atsushi Shimada, Hajime Nagahara, Rin-ichiro Taniguchi, Li He, “Image Annotation with Incomplete Labelling by Modelling Image Specific Structured Loss,” *IEEJ Transactions on Electrical and Electronic Engineering*, Vol.11, No.1, pp.73-82, 2015.

② Xing Xu, Atsushi Shimada, Hajime Nagahara, Rin-ichiro Taniguchi, “Learning multi-task local metrics for image annotation,” Multimedia Tools and Applications,” 2014. DOI: 10.1007/s11042-014-2402-7

③ Atsushi Shimada, Vincent Charvillat, Hajime Nagahara, Rin-ichiro Taniguchi, “Geolocation based Landmark Detection and Annotation -Towards Clickable Real World-,” *IEEJ Transactions on Electronics, Information and Systems*, Vol.133, No.1, pp.142-149, 2013.

〔学会発表〕（計　9 件）

① Xing Xu, Yang Yang, Atsushi Shimada, Rin-ichiro Taniguchi, Li He, “Semi-supervised Coupled Dictionary Learning for Cross-modal Retrieval in Internet Images and Texts,” ACM Multimedia 2015, pp.847-850, 2015.

② Xing Xu, Atsushi Shimada, Rin-ichiro Taniguchi, Li He, “Coupled Dictionary Learning and Feature Mapping for Cross-Modal Retrieval,” International Conference of Multimedia and Expo (ICME), 2015.

③ Hao Liu, Xu Xing, Hideaki Uchiyama, Atsushi Shimada, Hajime Nagahara, Rin-ichiro Taniguchi, “Query Expansion with Pairwise Learning in Object Retrieval Challenge,” 21th Korea-Japan Joint Workshop on Frontiers of Computer Vision, 2015.

④ Xing Xu, Atsushi Shimada, Rin-ichiro Taniguchi, “Exploring Image Specific Structured Loss for Image Annotation with Incomplete Labelling,” 12th Asian Conference on Computer Vision (ACCV2014), 2014.

⑤ Xing Xu, Atsushi Shimada, Rin-ichiro Taniguchi, “Tag Completion with Defective Tag Assignments via Image-Tag Re-weighting,” International Conference of Multimedia and Expo (ICME), 2014.

⑥ Takashi Ito, Atsushi Shimada, Hajime Nagahara, Rin-ichiro Taniguchi, “Selection of Representative Photos based on Sightseeing Behavior Analysis,” 20th Korea-Japan Joint Workshop on Frontiers of Computer Vision, 2014.

⑦ Xing Xu, Atsushi Shimada, Rin-ichiro Taniguchi, “Image Annotation by Learning Label-specific Distance Metrics,” 17th International Conference on Image Analysis and Processing (ICIAP), 2013.

⑧ Xing Xu, Atsushi Shimada, Rin-ichiro Taniguchi, “Latent Topic Model for Image Annotation by Modeling Topic Correlation,” International Conference of Multimedia and Expo (ICME), 2013.

⑨ Xing Xu, Atsushi Shimada, Rin-ichiro Taniguchi, “Correlated Topic Model for Image Annotation,” 19th Japan-Korea Joint Workshop on Frontiers of Computer Vision, 2013.

６．研究組織

(1)研究代表者

谷口倫一郎（TANIGUCHI, Rin-ichiro）

九州大学・大学院システム情報科学研究院・教授

研究者番号：２０１３６５５０

(2)研究分担者

長原一（NAGAHARA, Hajime）

九州大学・大学院システム情報科学研究院・准教授

研究者番号：８０３６２６４８

島田敬士（SHIMADA, Atsushi）

九州大学・基幹教育院・准教授

研究者番号：８０４５２８１１

内山英昭（UCHIYAMA, Hideaki）

九州大学・大学院システム情報科学研究院・助教

研究者番号：９０７３５８０４

(3)研究協力者

徐行（XU, Xing），九州大学大学院システム情報科学府博士課程（研究実施時）．