位置情報付き類似画像を用いた 未知画像のタグ推定における精度評価実験

和多田 吉樹† 鎌原 淳三‡ 長松 隆* 田中 直樹**

神戸大学大学院 海事科学研究科 〒658-0022 兵庫県神戸市東灘区深江南町 5 丁目 1-1

E-mail : †090w115w@stu.kobe-u.ac.jp , ‡kamahara@maritime.kobe-u.ac.jp

*nagamatu@kobe-u.ac.jp , **ntanaka@ maritime.kobe-u.ac.jp

あらまし 本研究の目的は、位置情報及びタグを有した画像データベースに対してタグと関連付けのない画像（以下、未知画像と呼ぶ）で問い合わせを行い、結果として画像の具体的な内容を表すタグを類似画像検索により求めるタグ推定システムの検索精度を向上させることである。このシステムでは位置情報に基づいてクラスタリングした画像群から特徴量に基づいて類似画像候補を選択し、候補画像のタグを導出する。本研究では、画像データセット CoPhIR のうち約 20 万枚を対象に、MPEG-7 の 5 つの特徴量を用いてそれぞれ実験を行い、精度を比較・評価した。また、計算時間が膨大なため、様々なサンプリングデータ数でも同様の実験を行った。

キーワード 類似画像検索、位置情報、タグ推定、画像特徴量

1 研究の背景と目的

1．1 研究の背景

近年、インターネット利用者の増加に加え、デジタルカメラなどの普及により個人でデジタル写真のデータを扱うことが容易になり、その結果 WWW(World Wide Web)上に莫大な数の画像が流通している。特に Flickr を代表とする、デジタルカメラなどによる写真を共有するコミュニティサイトでは 1 日に何百万枚というデジタル写真が世界各地のユーザから投稿されている。個人が所有するコンピュータの性能の向上、及び通信インフラの拡充にともない、画像データは文書データなどにも劣らず容易に流通・蓄積可能な情報になったと言える。その結果、ネットワーク上に莫大な数の画像が拡散していることから、大量の画像の中から目的の画像を見つけることは、課題の一つとなっている。

現在、単語で検索を行い検索結果が画像で表示されるものは Google[1]の画像検索をはじめ一般化している。つまり、画像が単語情報を有していれば、単語からその単語に関する画像を得ることは技術的には、文書検索と同じである。一方、画像検索において、画像から類似画像や単語情報を得るといった検索手法も、近年 Google Search by Image を用いて行えるようになった。しかし、その精度は明らかではなく、単語情報を求めようとする画像が、既存画像と同一性がない新規の画像で一切タグ等のテキスト情報を持たない場合は難しい。また、言葉と画像の関係付けが問題となり、求めていない検索結果が表示されることもしばしばあり、依然困難な課題となっている。

未知画像から、その画像の単語情報を導き出す手法は大きく分けると二つあり、画像認識により内容理解を行うものと、類似画像のメタデータを使うものがある。今回は、後者のメタデータとしてのタグ利用に焦点をあてて研究を行った。

検索結果に具体的な情報である単語が含まれるデータとして、写真を共有するコミュニティサイトである Flickr[2]のタグに含まれている情報に着目した。Flickr のタグとは、Flickr に登録したユーザが、Flickr 内の写真に対して分類する際に自由に付けることができるキーワードのことであり、メタデータの一種である。Flickr 内の写真は自由にタグを付けることが許可されており、撮影場所・撮影対象などの具体的な情報を表すタグが多い。このことから、Flickr のタグを検索結果の単語の情報として利用すると、検索質問である画像の具体的な単語を得ることが期待できる。

本研究では、Flickr の画像から成るデータセットとして CoPhIR を用い、画像のタグ情報を利用して、未知の画像から具体的な単語を推定する手法を提案し、評価を行った。

1．2 CoPhIR

本研究はデータセットとして CoPhIR[3]を用いる。CoPhIR （ Content-based Photo Image Retrieval Test-Collection）とは、研究に利用可能な世界最大のマルチメディアメタデータの集まりで、約 1 億 600 万枚の画像から成る。CoPhIR のデータセットは 1 億 600 万枚の Flickr からの画像のタイトル・撮影場所・タグ・コメントなどのタグ情報、そして、5 つの MPEG-7 特徴量（Scalable Color,Color Structure,Color Layout,Edge Histogram,Homogeneous Texture）を XML 形式のファイルに格納している。1 枚の画像あたり平均「3.1 個の

タグ」、「42 回の閲覧数」、「0.53 のコメント」がある。また、CoPhIR のデータセットのうち位置情報を持った画像は 8,655,289 枚（全体の 8.17%）である。

１．３　関連研究

位置情報が付与された画像を対象とした既存の研究の例を以下に挙げる。

帆足ら[4]は、検索対象画像に付与されている位置情報に基づくクラスタリングを行い、有意なクラスタを抽出してから、content-based クラスタリングを行う手法を提案している。この研究は、単語から画像を検索することを前提とした研究であり、画像から単語情報を導出する本研究とは目的が異なる。

Naaman ら[5]の研究では、写真に付与された位置情報・撮影時間に基づき、個人の写真コレクションに自動的に名前を付けて整理する手法を提案している。この研究は写真データを整理することが目的であるため、名前付けは地理名称に留まる。また、基本的に個人の写真コレクションが分析対象となっており、評価実験で使用した画像も小規模である。

Kennedy ら[6]は、Flickr の位置情報付き画像を利用し、位置情報・タグ・画像特徴量を用いてランドマークを表すタグを推薦する方法を提案している。この研究は、データセットはサンフランシスコ内の画像に限定されており、また既にタグ情報が付いている画像により相応しいタグを推薦することを目的としている。

Abbasi ら[7]の研究では、本研究と同じく CoPhIR を用いて位置情報・MPEG-7 特徴量（Edge Histogram・Color Layout）・タグといった各々の特徴量でクラスタリングを行い、その各モデルから出力されるタグの精度を検証している。この研究で最も良い結果が得られているのは、位置情報によるクラスタリングを利用してタグを推薦する方法であるが、この方法だと位置情報が近く同じクラスタに属す画像であれば、全て同じタグが推薦されてしまう。

島田ら[8]は、画像特徴量から単語情報を推定と、位置情報から単語情報の推定を行い、これら２つの結果から最終的に単語情報を推薦する手法を提案している。画像特徴量のみから単語情報を推定する場合、その画像特徴量を含む画像の単語情報は全て推薦対象となってしまうため、大まかな単語情報が推定される場合が多いと考えられる（「海」、「山」、「タワー」など）。また、２つの推定方法を同等に評価しているため、この方法では結果的に大まかな単語情報が推定されると考えられる。また、実験は Flickr から収集した 87 のシーンに留まる。

Papadopoulos ら[9]の研究では、タグ付けされた写真からランドマークやイベントを自動的に検出するための新たな手法を提案している。この手法は、まず視覚的類似度とタグ類似度のグラフを作成し、それらからハイブリッド画像類似度グラフを生成する。これら３つのグラフを用いて、グラフベースの画像クラスタリングを行う。その後、それぞれのクラスタを「Landmark」、もしくは「Event」に分類する。そして、Landmark クラスタに分類されたクラスタは位置情報をもとに、更に分類・ラベリングされる。しかし、この手法はクラスタが長い期間をまたがる複数の類似したイベント（例えば、結婚式など）構成の場合等に誤分類されるケースがある。

１．４　本研究の目的

本研究は、単語との関連付けのない未知画像で問い合わせを行い、結果として画像の具体的な内容を表す単語を求めるシステムを作ることを目標とする。

近年、位置情報が記憶されるカメラ機能がついた携帯電話などの普及により、位置情報が付与されている画像が増加してきたことに注目し、その画像の位置情報と画像特徴量を手がかりに、単語情報が全く付いていない画像から具体的な単語情報を推定する。比較的大きなランドマーク等であれば、位置情報と地図のデータから抽出できる可能性はあるが、位置情報だけでは実際にはそのランドマークが写っていない画像に対しても、ランドマークの名称を提示してしまうことになり、画像から単語を抽出したことにはならない。そこで、位置情報に加え MPEG-7 特徴量の中から５つの特徴量を用いることで、未知画像の単語情報を推定する。

それにより得られるタグ情報と、未知画像の具体的な内容を表しているであろう単語情報が、どれくらい類似しているのか、５つの MPEG-7 特徴量における精度を評価することを目的としている。

２ タグ情報導出システム

２．１ システム概要

本研究で用いたタグ情報導出システム[10]は、予めデータセットの位置情報を基にクラスタリングを行い、モデルを作成する。未知画像からタグ情報を推定する際は、未知画像の位置情報に最も距離が近いクラスタをモデルから探し、そのクラスタ内で画像特徴量が近い画像群のタグ集合から、未知画像の内容を表す単語を抽出する。図 1 に本システムの流れを示す。

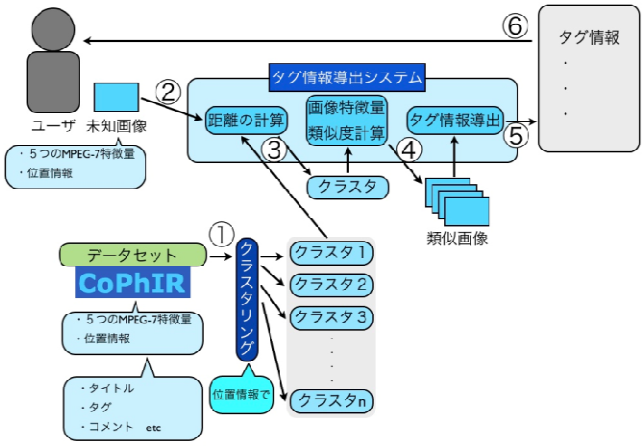

図１．タグ情報導出システムの流れ

①データセットの位置情報をもとにあらかじめクラスタリングを行っておく
②未知画像を入力する
③未知画像の位置情報と最も距離が近いクラスタを求める
④そのクラスタ内にある画像特徴量が類似した画像を求める
⑤類似した各画像のタグ情報を集める
⑥検索結果としてタグ情報をフィードバックする

２．２　タグ情報の導出

画像特徴量によって未知画像の類似画像を求めた後に、求められた複数の画像の代表となるタグを推薦する。そして、類似画像の代表しているタグを未知画像の単語情報として導出する。導出すべきタグを特定するために、求められた類似画像のタグの重複した数を数える。重複した数をタグの推薦数とすると、推薦数が多いタグは、より多くの類似画像がそのタグを使用しており、また逆に推薦数が少ないタグは一部の類似画像しかそのタグを使用していないことになる。タグ情報導出システムではより推薦数が多いタグを未知画像の単語情報として導出する。

データセットのタグの全集合を T、そのタグの総数を K としたとき、

$$T = \{t_1, t_2, \ldots\ldots, t_K\}$$

である。

クラスタ全体を C、クラスタの数を m としたとき、

$$C = \{C_1, C_2, \ldots\ldots, C_m\}$$

である。

未知画像 I_X と位置情報が近いクラスタを C_S としたとき、

$$C_S' = \{I_{S1}, I_{S2}, \ldots\ldots, I_{Se} \mid sim_Y(I_i, I_X) > threshold\}$$

である。ここで、$sim_Y(I_i, I_X)$ は、特徴量 Y による画像間のコサイン類似度である。また *threshold* はコサイン類似度に対する閾値であり、任意に指定できる。

C_S' は C_S の中で画像 I_X との類似度が高い画像の集合であり e はその画像の総数である。この時、C_S' 中の画像 $I_{S1}, \ldots, I_{Se}$ は、それぞれタグ $t_{I_{Sk1}}, \ldots, t_{I_{Skn}}$ を持つ。よって、以下のように表せる。

$$I_{S1} = \{t_{I_{S11}}, t_{I_{S12}}, \ldots\ldots, t_{I_{S1p}}\}$$

$$I_{S2} = \{t_{I_{S21}}, t_{I_{S22}}, \ldots\ldots, t_{I_{S2r}}\}$$

$$\vdots$$

$$I_{Sk} = \{t_{I_{Sk1}}, t_{I_{Sk2}}, \ldots\ldots, t_{I_{Skn}}\}$$

$$\vdots$$

$$I_{Se} = \{t_{I_{Se1}}, t_{I_{Se2}}, \ldots\ldots, t_{I_{Seu}}\}$$

推薦されるタグ $Num(t, C_S')$ は以下のようになる。

$$Num(t, C_S') = \sum_{i=1}^{u} \sum_{j=i+1}^{u} count_t(I_i, I_j)$$

ここで、$count_t$ はタグ t が重複した数であり、I_i、I_j に共に存在するときは１（ただし $i \neq j$）、それ以外は０となる。

また、最終的に導出されるタグ T_R は、

$$T_R = \{Num(t, C_S')の上位N個\}$$

である。

本システムは推薦されたタグの数の上位何位までを導出するか設定できる。ただし、最終位が複数ある場合、それら全てを導出する。図２を例にすると、上位２位までを導出するとすれば、導出数は２つだが、上位３位までを導出する場合は導出数は５つになる。

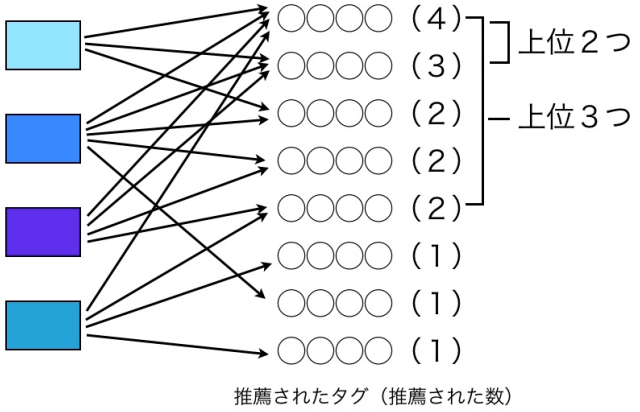

図 2. タグの重複

3 精度評価実験

3．1 実験方法

実験は、未知画像をタグ情報導出システムに入力したとき、どれだけ正しい単語情報が出力されるかを評価する。これは、データセットをタグ情報を持たないと仮定するテストデータ、及びタグ情報導出システムのデータベースとして機能するトレーニングデータに分けることによって、システムの出力結果がテストデータセットで元々付与されていたタグと一致する数を評価することにより行う。

本実験は、用いる画像特徴量によって以下の 5 種類のパターンによって行った。

①Scalable Color のコサイン類似度が x 以上のタグを推薦
②Edge Histogram のコサイン類似度が x 以上のタグを推薦
③Color Structure のコサイン類似度が x 以上のタグを推薦
④Color Layout のコサイン類似度が x 以上のタグを推薦
⑤Homogeneous Texture のコサイン類似度が x 以上のタグを推薦

本実験では、x=0.5,0.6,0.7,0.8,0.9 の 5 パターンを行った。また、各実験において 5-fold cross validation とするため、全部で 125 パターンの実験を行った。

実験では推薦された数が多い上位 10 個のタグをタグ情報導出結果とした。これは、Scalable Color のコサイン類似度が 0.7 以上のタグを全て導出する実験を行ったところ、約 250 個ものタグが導出されてしまい、あまりに多くのタグを導出すると未知画像と関係のない単語が多く含まれてしまうためである。

各実験につき、
・タグが導出されなかった画像の枚数（sim miss）
・導出されたタグの数（length）
・テストデータの本来もっていたタグと導出したタグが一致した数（count out）
・テストデータのタグがトレーニングデータ全体の中にある数（count all）
を記録し、これらの値からシステムを評価する。

3．2 *k*-fold cross validation

実験結果としてより信頼できる結果を得るため、*k*-fold cross validation により実験を行った。*k*-fold cross validation とは、標本群を *k* 個に分割して、そのうち 1 つをテスト事例とし、残る(*k*－1)個をトレーニング事例とする方法である。*k* 個に分割された標本群それぞれのテスト事例で *k* 回計算を行い、最終的に *k* 回の結果を平均して 1 つの推定結果を得る。

本実験では *k*=5 として、5-fold cross validation を行った。データセットを 5 分割し、そのうち 1 個をテストデータ､残る 4 個をトレーニングデータとして､各々のパターンを 5 回計算し、その結果の平均を実験結果とする。

3．3 データセット

本研究において、タグ情報導出を行うために用いる検索対象の画像の集合をデータセットと定義する。

今回使用したデータセットは、約 1 億 600 万枚の画像から成る CoPhIR の内の位置情報付きの画像データである 8,655,289 枚（全体の約 8.17%）の中から任意に取り出した、209,095 枚（CoPhIR の位置情報つき画像の約 2.4%（全体の約 0.2%））の画像データで構成される。これらの画像は全て位置情報とタグ情報を持っており、世界の主要都市を多くカバーしている。

3．4 テストデータ

システムの精度を測るために、単語情報を全く持たない位置情報のみを持った未知画像が必要である。データセットの一部を、単語情報を全く持たない画像と仮定して、これらをテストデータとする。テストデータは 209,095/5 の 41,819 枚となる。

このテストデータ数が多いと実験に多大な時間がかかってしまうため、本実験では、41,819 枚の中からランダムに 3,000 枚をサンプリングし、その 3,000 枚をテストデータとした実験も行った。

テストデータ数を少なくする際に、少なくしすぎると実験の精度が落ちてしまうため、どの程度までテストデータを減らしても精度が変わらないかを考察するために、Scalable Color・Edge Histogram・Color Structure・Color Layout の 4 つの特徴量を用いて、いく通りかのテストデータ数で実験を行った。図 3・4 にその結果を示す。

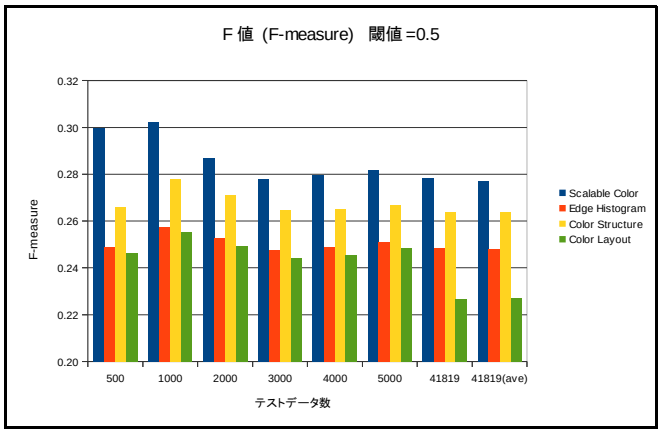

図 3.　F 値（閾値=0.5）

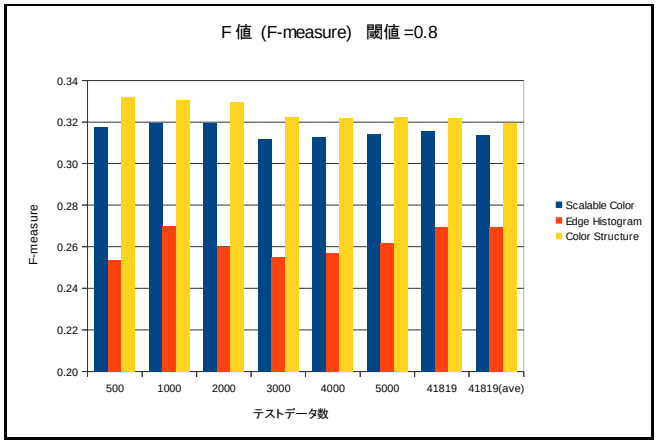

図 4.　F 値（閾値=0.8）

テストデータ数を決める際に行ったこの実験は、閾値が 0.5 と 0.8 で、5 分割のうちの 1 セットである test1 に関してのみ行った（5-fold cross validation を行っていない）。上記のグラフより、テストデータ数を 3,000 枚より少なくしてしまうと、本来のテストデータ数（41,819 枚）を用いて行った実験から得られた F 値と大きくかけ離れてしまうため、今回の実験はテストデータ数を 3,000 枚で行うことにした。

３．５　トレーニングデータ

タグ情報・位置情報ともに持っている画像データをトレーニングデータとする。本実験ではこのトレーニングデータがタグ情報導出システムのデータベースとして機能する。トレーニングデータは 209,095－41,819 の 167,276 枚となる。なお、実験時間を短縮するためにテストデータ数を 3,000 枚と少なくしたが、実験の精度を下げないため、トレーニングデータの数は 167,276 枚そのままを使用している。

３．６　評価方法

タグ情報導出システムの性能の評価を適合率・再現率・F 値によって行う。適合率は、検索結果として導出されたタグ情報に、どれだけ元のテストデータが持っているタグを含んでいるかという正確性の指標である。再現率は、検索対象としているトレーニングデータの中で検索結果として適合しているタグのうち、どれだけそのタグを導出できているかという網羅性の指標である。適合率（precision）・再現率（recall）は以下の式で求められる。

$$precision = \frac{R}{N} \qquad recall = \frac{R}{C}$$

R:テストデータが本来持っていたタグと導出したタグが一致した数
N:導出されたタグの数
C:テストデータのタグがトレーニングデータ全体の中にある数

また、適合率を上げれば再現率が下がり、再現率を上げれば適合率が下がる傾向にあるため、最終的な精度評価指標として F 値（F-measure）を用いる。F 値は適合率と再現率の調和平均であり、

$$F-measure = \frac{2 \times precision \times recall}{precision + recall}$$

によって求められる。F 値が高ければ、システムの性能が良いことを意味する。

４　実験結果

４．１　約２０万枚の画像を用いて行った実験結果

テストデータ数を減らさずに行ったこちらの実験では、Scalable Color・Edge Histogram・Color Structure・Color Layout・Homogeneous Texture の 5 つの画像特徴量において、閾値が 0.5～0.9 まで全ての実験を行った。

本システムの評価を表す指標である、適合率・再現率・F 値の実験結果を図 5～7 に示す。

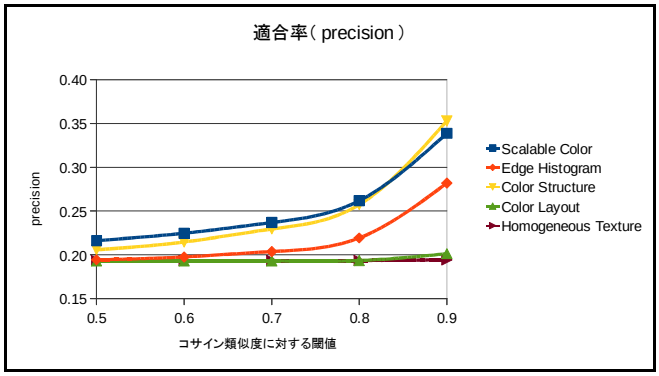

図 5. 適合率（テストデータ数=41,819 枚）

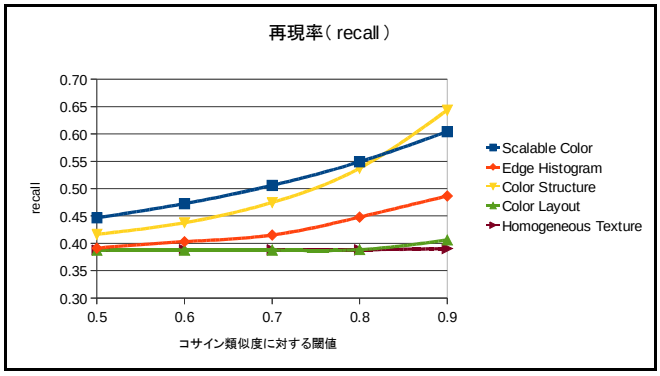

図 6. 再現率（テストデータ数=41,819 枚）

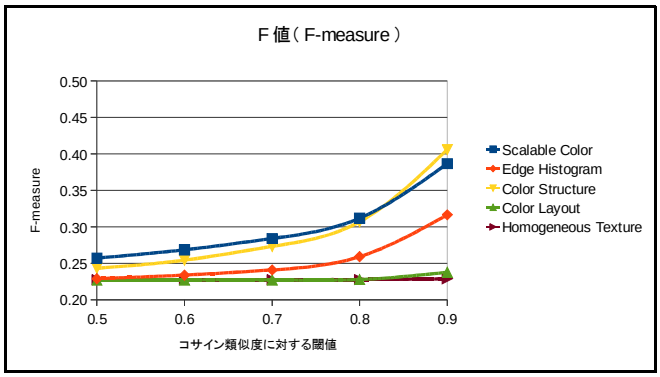

図 7. F 値（テストデータ数=41,819 枚）

本実験のプログラムは主に Perl を用いて計算している。実験に要した時間は、実験に用いる PC の性能や、CPU・メモリの使用状況、また、使用する特徴量によっても異なるが、閾値=0.5 の時で約 37 日、閾値=0.8 の時で約 10 日程であった。今回は複数の PC を用いて並列に実験を行った。なお、現在ではプログラムのアルゴリズムを修正したことにより、プログラムの計算時間を大幅に短縮することができた。上述したように、様々な要因によって計算時間は多少前後するものの、数日（長い時で 30 日以上）かかっていた計算を、約 3～4 時間程度にまで短縮することができた。所要時間に関しては、コサイン類似度に対する閾値を上げる程、候補画像の枚数が減るので、閾値が低い時に比べて計算時間が短くなる。

4．2 3,000 枚のテストデータ数で行った実験結果

テストデータ数を 3,000 枚として行ったこちらの実験でも、5 つの画像特徴量それぞれにおいて、閾値が 0.5～0.9 の全ての実験を行った。また、こちらの実験では、閾値が 0.8～0.9 の部分にかけて、適合率や最終的な精度を示す F 値の順番が入れ替わる部分が生じたため、0.8 と 0.9 の中間である、閾値 0.85 に関しても同様の実験を行った。

4.1 の実験結果と同様に、適合率・再現率・F 値の実験結果を図 8～10 に示す。

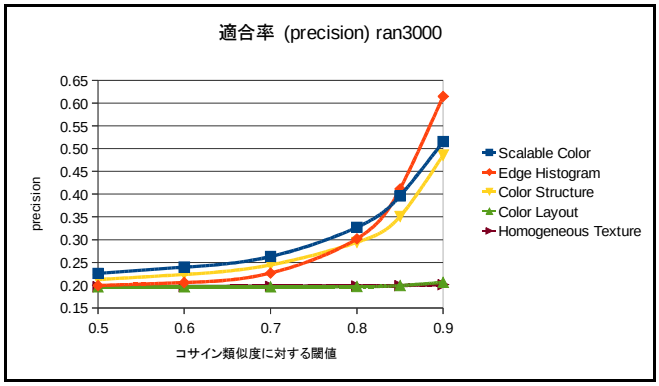

図 8. 適合率（テストデータ数=3,000 枚）

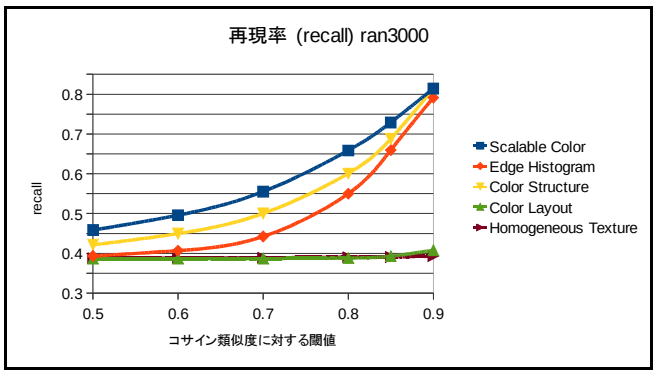

図 9. 再現率（テストデータ数=3,000 枚）

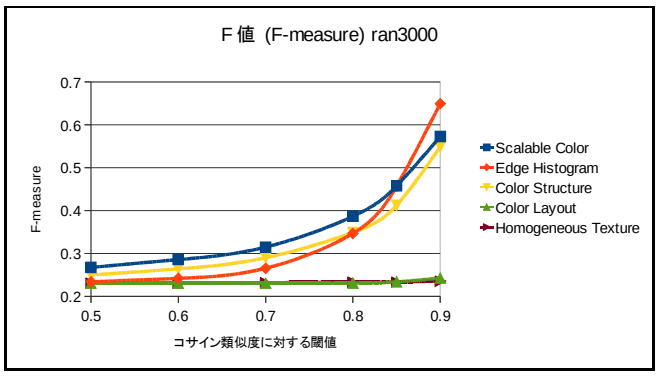

図 10. F 値（テストデータ数=3,000 枚）

実験に要した時間は、閾値=0.5 の時で約 48 時間、閾値=0.8 の時で約 6 時間程であった。（実験プログラムのアルゴリズム修正前の所要時間）

5 考察

実験結果より、位置情報によって予めクラスタリングを行うことで、画像特徴量を使ったタグ情報導出を効果的に行うことができた。

Scalable Color・Color Structure に関しては、テストデータの本来もっていたタグと導出したタグの一致した数（count out）と、テストデータのタグがトレーニングデータ全体の中にある数（count all）との値は非

常に近く、高い再現率が得られた。再現率に比べ、適合率がそれほど高くなかったため、F 値はそれほど良い結果が得られなかったが、本研究で用いた５つの画像特徴量の中では比較的良い結果が得られた。この２つの特徴量に関しては、画像中の色に関する特徴量であり、色の配置は関係しない。つまり、撮影対象物が同じであれば、撮影のアングルが違っても画像特徴量の類似度が高い可能性があると推測できる。よって、この２つの特徴量に関しては比較的良い結果が得られたと考えられる。

Edge Histogram に関しては、上述した２つの特徴量ほど良い結果は得られなかった。

Color Layout・Homogeneous Texture に関しては、適合率・再現率ともにとても低く、結果として F 値はとても低い値となった。また、この２つの特徴量に関しては、コサイン類似度に対する閾値を変化させても、適合率･再現率ともにほとんど変化が見られなかった。

Edge Histogram・Homogeneous Texture に関しては、画像中の模様に関する特徴量であり色は関係ない。つまり、撮影のアングル等が違うと画像特徴量が異なってくると推測される。よって、この２つの特徴量に関しては良い結果が得られなかったと考えられる。

Color Layout は画像中の色配置を表す特徴量である。つまり、撮影対象物が写真のどこに位置するか、どれくらいの大きさで写っているのかによって画像特徴量が異なってくる。よって、この特徴量に関しても良い結果が得られなかったと考えられる。

以上より、類似画像を検索する過程において用いる画像特徴量として、Scalable Color と Color Structure は有意義であると考えられるが、Edge Histogram ・Color Layout・Homogeneous Texture を用いるのは、有意義ではないと考えられる。

６ まとめ

今回は時間の都合上、テストデータ数を減らして実験を行ったが、本来のテストデータ数（41,819 枚）で行った実験の F 値（図 7）と、テストデータ数を減らして（3,000 枚）行った実験の F 値（図 10）を比較すると、3,000 枚で行った実験の方が、コサイン類似度に対する閾値を高くする程、F 値の上がり具合が大きくなるという結果になった。テストデータ数が多い程（今回の場合は 41,819 枚で行った実験の方）実験結果としては信頼性が高いと考えられるため、テストデータ数を 3,000 枚にまで少なくして行った実験に関しては、閾値の高い部分においてバラつきが大きいと考えられる。また、５つの画像特徴量ごとの傾向を、実験結果より確認することができた。

本研究では検索対象画像として、位置情報を持っているランドマークや建築物、及び風景など移動しないオブジェクトの画像を想定しており、Papadopoulos らの研究において「Landmark」とは別の「Event」に分類される人物、車、商品などの移動するオブジェクトには不向きである。

今後の展望としては Papadopoulos らの提案する手法を用いて､データセットの画像を｢Landmark｣と｢Event｣に分類した後に、｢Landmark｣ と分類された画像に対してのみ、本システムを用いて類似画像からタグ情報導出を行うといった手法を検討していきたい。

また、本研究では CoPhIR の画像データセットを用いており、画像のタグ情報の中には、撮影日時・ユーザ名・カメラメーカ等のように、画像の具体的な内容を表すタグ情報として相応しくないタグも含んでいる。実際に本実験では、タグ情報導出結果として、このような画像の具体的な内容を表す単語情報として相応しくないタグも導出されてしまっている。これらの不必要なタグ情報をあらかじめ取り除いておくことも今後の検討課題である。

謝辞

本研究は科研費（ 22300035 ）の助成を受けたものである。

参考文献

[1] http://images.google.co.jp/

[2] http://www.flickr.com/

[3] http://cophir.isti.cnr.it/

[4] 帆足 啓一郎, ルンドステット マグナス, 上向俊晃, 松本 一則, 滝嶋 康弘, “位置情報メタデータを利用した画像検索手法の実装と評価,” 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解 108(94), pp.65-70 (2008)

[5] Naaman Mor, Song Yee Jiun, Paepcke Andreas, Garcia Molina Hector, “Automatic Organization for Digital Photographs with Geographic Coordinates,” Fourth ACM/IEEE Joint Conference on Digital Libraries (2004)

[6] Lyndon S. Kennedy, Mor Naaman, “Generating diverse and representative image search results for landmarks,” In WWW '08: Proceeding of the 17th international conference on World Wide Web (2008), pp. 297-306 (2008)

[7] Rabeeh Abbasi, Marcin Grzegorzek, and Steffen Staab, “Large Scale Tag Recommendation Using Different Image Representations,” Semantic Multimedia, Vol. 5887Springer Berlin / Heidelberg (2009), pp. 65—76 (2009)

[8] 島田敦士, Vincent CHARVILLAT, 長原一, 谷口倫一郎, “撮影位置情報を利用した画像アノテーションに関する検討,” 電子情報通信学技報, vol. 110, no. 296, PRMU2010-113, pp. 1-6 (2010)
[9] Symeon Papadopoulos, Christos Zigkolis, Yiannis Kompatsiaris, Athena Vakali, “Cluster-Based Landmark and Event Detection for Tagged Photo Collections”, IEEE Multimedia, January-March 2011 (vol. 18 no. 1) pp. 52-63
[10] 岩井俊英, 鎌原淳三, 長松隆, "位置情報付き画像クラスタリングによる未知画像からの単語情報の導出", 電子情報通信学会 総合大会, 学生ポスターセッション, ISS-P-132 (2011)