タグの共起と画像の類似性に基づくタグ付け支援システム

瀬崎 直人*　竹内 謹治　黄瀬 浩一
大阪府立大学大学院工学研究科

Tagging System Based on Co-occurrence of Tags and Similarity of Images

Naoto Sezaki*　Kinji Takeuchi　Koichi Kise
Graduate School of Engineering, Osaka Prefecture University

Abstract　This report presents a method for recommending tags of images.The proposed method presents to the user various tags with high precision by taking into account both the co-occurrence of tags and the similarity of images. Additional search with the user feedback enables us to find some new tags relevant to the image of interest.In order to test the proposed method, we compare it with a method based only on the co-occurrence of tags, as well as a method based only on the similarity of images. From the experimental results using 5377 images, we have confirmed that the proposed method is capable of finding more tags as compared to the methods for comparison. We have also confirmed that the additional search is effective to find additional tags.

キーワード　画像検索　タグ　共起　LSI　類似画像　EMD

Key words　image retrieval　tag　co-occurrence　LSI　similar image　EMD

1　はじめに

今日，デジタルカメラやインターネットの普及により，Web ページ上には大量の画像が存在している．これらの大量の画像に対して，ユーザが必要とする画像を得る一手法として，キーワードから画像を検索する手法がある．これを実現するためには，画像に対して予めキーワードを付与しておく必要がある．このような問題を解決する一手法として，folksonomy [1] という手法が提案されている．この手法では，個々のユーザが自分の視点でデータにタグと呼ばれるキーワードを付与し，それらのタグを共有することにより，各ユーザはタグを通してそれぞれのデータにアクセスする．この手法は，flickr [2] を代表とする画像共有サイトなどで応用され，注目を集めている．画像共有サイトでは，ユーザが画像に対してタグを付与することにより索引付け (タグ付け) を行う．

しかし，初心者はどのようなタグをつければいいのか分からないため，初心者がタグ付けを行った画像は十分なタグが付与されていないという問題点がある．一方で，熟練したユーザによって十分にタグが付与された画像も多く存在する．このような画像のデータベースにおいて，互いに関連する語は，同時に同じ画像のタグになっている (共起している) ことが多い．そのため，ユーザが入力したタグと共起することが多いタグを画像の関連語として提示できると考えられる．また，類似している画像は同じタグを含んでいることが多いので，これらのタグも関連語として提示できると考えられる．

そこで本研究では，この考えを利用し，タグの共起と画像の類似性に基づくタグ付け支援システムを提案する．本手法の特徴は，タグの共起だけでなく類似画像の関連語も抽出することで，多様で精度の高い関連語を新しいタグの候補としてユーザに提示できる点である．また，ユーザが新たに付与したタグを入力のタグに含め，関連語の再検索を行うことで，さらに新しい関連語を抽出することができる．タグの共起のみを用いる手法や類似画像のみを用いる手法と比較実験を行った結果，提案手法はより多くの有効な関連語を抽出できることがわかった．また，これらの有効なタグを入力タグに追加し，関連語の再検索を行うことにより，新たに有効な関連語を抽出できることも明らかになった．

2　関連研究

画像検索には，画像から抽出した色・テクスチャなどの画像の内容に基づく画像検索 (Content-Based Image Retrieval:CBIR) [3] と画像に与えられたキーワードに基づく画像検索 (Text-Based Image Retrieval:TBIR) が存在する．

CBIR では画像の色や形状などを特徴量として類似画像を

検索する．このような特徴量では画像の内容を的確に表すことが難しいので，満足のいく検索性能が得られていない．この問題を解決するため，Rui らは，関連フィードバックを用いて検索性能の向上を行う手法を提案している [4]．

TBIR では，既存の情報検索の技術を用いて画像検索を行うことが可能である．しかし，TBIR を実現するためにはデータベース中の画像に対して，テキストによる索引付けを行う必要がある．Web ページを対象とした自動索引付け手法がいくつか提案されている [5, 6, 7, 8]．しかし，Web ページは記述のされ方も様々であるため，Web 上の画像に対して安定して充実した索引付けを行うことは困難である．

一方，Web ページ中には，folksonomy [1] という手法を用いてブックマークや画像を分類するサービスが存在する．この手法では，サービスを利用するユーザがデータにタグと呼ばれるメタデータを付与する．そして，付与されたタグを共有することにより，各ユーザはタグを通してそれぞれのデータにアクセスする．flickr はタグ付けを行う対象を画像としたサービスである．画像にタグ付けを行うことで，あるタグからの繋がりにより，新しい画像を発見することが可能となる．

3 タグの共起と類似画像を用いたタグ付け支援システム

画像共有サイトには，熟練したユーザによって十分にタグが付与された画像が多く存在する．提案するシステムでは，このような十分にタグ付けされた画像を利用して，新たに画像の関連語を抽出し，ユーザに提示する．

3.1 システムの概要

タグ付け支援システムの全体構成を図 1 に示す．提案システムでは，画像と画像に付与された少数のタグのペアを入力とする．付与するタグはユーザが自由に設定する．

この画像とタグのペアを入力とし，「関連語抽出モジュール」によって，次にタグ付けするべき語を検索し，ユーザに提示する．このモジュールでは，タグが十分に付与された画像のベータベースにアクセスし，関連語を抽出する．関連語の抽出は，入力したタグとベータベース中にある画像のタグの共起，また類似画像のタグを利用して行う．この処理により得られた関連語を新たなタグの候補として出力し，ユーザはそれを画像の新たなタグとして付与するかどうか判断する．この際，タグの候補を出力すると同時に，それらが索引付けられている類似画像も出力する．これにより，ユーザは類似画像から視覚的にタグの候補を付与するかどうか判断することができる．

さらに，このシステムでは，ユーザが新たに付与したタグを入力のタグに含め，関連語の再検索を行う．入力とするタグの数を増やし再検索を行うことで，新しい関連語をユーザに提示する．これにより，ユーザはより多くのタグを画像に付与することが可能となる．

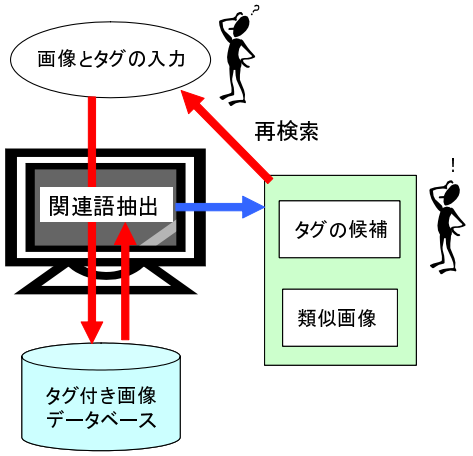

図 1　システムの概要

3.2 タグの共起を利用した関連語抽出

互いに関連する語は，同時に同じ画像のタグになっている (共起している) ことが多いと考えられる．この考えを利用し，入力したタグと共起することが多いタグを画像の関連語として抽出する．具体的な手順は以下の通りである．

まず，画像に付与されているタグをタグベクトルとして表現する．データベース中に α 種類のタグ $w_1, w_2 \cdots w_\alpha$ が存在するとき，画像 I_i を表すタグベクトル $\boldsymbol{t}_i$ を次のように表現する．

$$\boldsymbol{t}_i = [t_{i1}, t_{i2} \cdots t_{i\alpha}]^t \tag{1}$$

ここで t_{ij} は w_j の画像 I_i における重みである．タグ w_j が画像 I_i に付与されている場合は t_{ij} を 1，付与されていない場合は 0 とする．

画像 I_i, I_j のタグベクトル $\boldsymbol{t}_i, \boldsymbol{t}_j$ が与えられたとき，これらの間の類似度 $\mathrm{Sim}(\boldsymbol{t}_i, \boldsymbol{t}_j)$ は LSI 法 [9] を用いて計算する．具体的には，画像データベース中の各画像のタグベクトルをまとめた行列を

$$D = [\boldsymbol{t}_1, \boldsymbol{t}_2 \cdots \boldsymbol{t}_\beta] \tag{2}$$

とする．このとき D を以下のように特異値分解する．

$$D = U \Sigma V^T \tag{3}$$

この処理で得られた U の最初の k 次元の左特異値ベクトルのみから構成される行列を U_k とする．このとき $\boldsymbol{t}_i$ の k 次元表現 $\boldsymbol{t}^{(k)}$ は以下の式で与えられる．

$$\boldsymbol{t}_i^{(k)} = U_k^T \boldsymbol{t}_i \tag{4}$$

これを用いて，$\mathrm{Sim}(\boldsymbol{t}_i, \boldsymbol{t}_j)$ を以下の式で計算する．

$$\mathrm{Sim}(\boldsymbol{t}_i, \boldsymbol{t}_j) = \cos(\boldsymbol{t}_i^{(k)}, \boldsymbol{t}_j^{(k)}) = \frac{\boldsymbol{t}_i^{(k)} \cdot \boldsymbol{t}_j^{(k)}}{||\boldsymbol{t}_i^{(k)}||\ ||\boldsymbol{t}_j^{(k)}||} \tag{5}$$

提案システムでは，入力とする画像とデータベース中のすべての画像に対してタグベクトルの類似度を計算する．これらの画像を類似度が高い (付与するタグが類似している) 順

にソートし，上位 a_1 個の画像において，付与されているタグの出現回数を計算する．これらのタグを出現回数が多い順にソートし，上位 b_1 個のタグを関連語とみなす．

3.3 類似画像を利用した関連語抽出

提案システムでは，類似画像には同じようなタグが付与されていることを仮定する．例えば，同じ人物の画像であれば，そのタグとして同じ名前が付与されていると考えられる．この仮定のもと，類似画像に付与されているタグを，入力画像の関連語として抽出する．具体的な手順は以下の通りである．

提案システムでは画像の色の分布を用いて，類似画像を検索する．色の分布を表す画像の特徴量として color signature を用いる．color signature は色を表す代表色ベクトル $\boldsymbol{p}_i$ と色の画素数の割合 r_{p_i} の組 $(\boldsymbol{p}_i, r_{p_i})$ で表される．color signature 同士の距離を求めるには，Earth Movers Distance(EMD) [10] を用いる．color signature $P = \{(\boldsymbol{p}_1, r_{p_1}), (\boldsymbol{p}_2, r_{p_2}), \cdots (\boldsymbol{p}_m, r_{p_m})\}$，$Q = \{(\boldsymbol{q}_1, r_{q_1}), (\boldsymbol{q}_2, r_{q_2}), \cdots (\boldsymbol{q}_n, r_{q_n})\}$ と color signature の要素間の距離 $d_{ij} = d(\boldsymbol{p}_i, \boldsymbol{q}_j), (1 \leq i \leq m, 1 \leq j \leq n)$ が与えられたとき，P，Q 間の EMD は次のように表現される．

$$\mathrm{EMD}(P, Q) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} d_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij}} \tag{6}$$

ただし，f_{ij}は式 (6) を以下の条件で最小化する最適化問題の解である．

$$f_{ij} \geq 0 \qquad 1 \leq i \leq m, 1 \leq j \leq n \tag{7}$$

$$f_{ij} \leq r_{p_i} \qquad 1 \leq i \leq m \tag{8}$$

$$f_{ij} \leq r_{q_j} \qquad 1 \leq j \leq n \tag{9}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} = \min(\sum_{i=1}^{m} r_{p_i}, \sum_{j=1}^{n} r_{q_j}) \tag{10}$$

本手法の場合，色空間として人間の感覚に近いと言われている CIE L*a*b*色空間を用い，代表色ベクトルの距離 $d(\boldsymbol{p}_i, \boldsymbol{q}_j)$ としてユークリッド距離を用いる．

提案システムでは， 3.2 により得られるタグの類似度が高い l 個の画像に対して EMD を測定する．すべての画像に対して EMD を測定すると計算時間が膨大となるため， 3.2 の処理を行うことで，測定する対象を絞り込む．これらの画像を EMD が小さい (色の分布が類似している) 順にソートし，上位 $a_2 (a_2 \geq a_1)$ 個の画像において，付与されているタグの出現回数を計算する．これらのタグを出現回数が多い順にソートし，上位 b_2 個のタグを関連語とみなす．

3.4 関連語の提示

タグの共起により抽出される関連語の集合を G_{tag}，類似画像により抽出される関連語の集合を G_{img} とすると，最終的にユーザに提示する関連語の集合 $G_{\cup}$，$G_{\cap}$ を次のように求める．

$$G_{\cup} = G_{\mathrm{tag}} \cup G_{\mathrm{img}} \tag{11}$$

$$G_{\cap} = G_{\mathrm{tag}} \cap G_{\mathrm{img}} \tag{12}$$

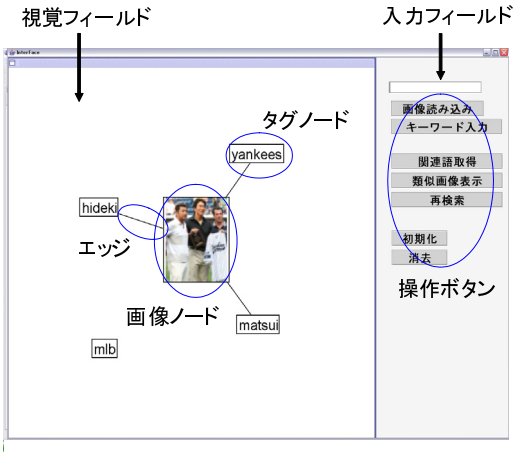

図 2　インターフェース

$G_{\cup}$ は， 3.2 と 3.3 の処理で抽出される，すべての関連語である．これにより，それぞれの処理だけで抽出する場合よりも，多くの関連語をユーザに提示する．また，$G_{\cap}$ はタグの共起と類似画像の両方から抽出される関連語なので，タグの候補の中でも特に関連度の高いタグとしてユーザに提示される．

3.5 関連語の再検索

提案システムでは，ユーザが新たに付与するタグを入力のタグに含め，関連語の再検索を行う．入力とする画像の言語的特徴量に，新たに付与するタグの重みを付け加え， 3.2, 3.3, 3.4 の処理を再び行い，新しい関連語をユーザに提示する．これにより，ユーザはより多くのタグを画像に付与することが可能となる．

3.6 インターフェースと実行例

提案システムのインターフェースを図 2 に示す．以下，ユーザがこのインターフェースを使って画像にタグを付与する手順を説明する．まず，ユーザは入力フィールドでタグ付けを行う画像とそれに付与するタグを入力する．これにより，視覚フィールドには，入力された画像とタグがノードとして出力され，ユーザによる操作が可能となる．このとき画像に付与されるタグノードとその画像ノードはエッジで結ばれている．次に，関連語抽出ボタンにより，「関連語抽出モジュール」を起動する．ユーザは，これにより出力されたタグの候補ノードから画像に付与するノードを選択する．タグの候補ノードは属する関連語の集合 ($G_{\cup}$ か $G_{\cap}$) ごとに色分けして出力される．また，類似画像表示ボタンで，関連語抽出に用いた類似画像のノードを出力することができる．ユーザは再検索ボタンにより，再び「関連語モジュール」を起動することにより，新たなタグの候補ノードが画像に出力される．

提案システムの実行例を図 3 に示す．ユーザは，入力する画像に対して，少数のタグを付与して関連語抽出を行う．その結果，視覚フィールド上に複数のタグの候補ノードが出力される．また，同時に関連語抽出に用いた類似画像も出力さ

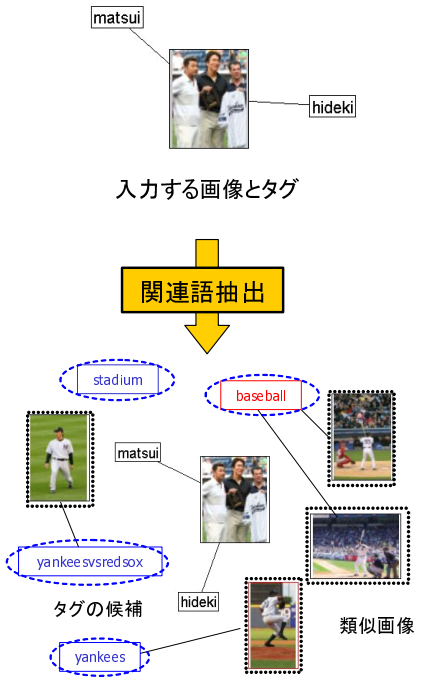

図 3　システムの実行例

表 1　実験に用いたパラメータ

k	a_1	b_1	l	a_2	b_2
500	20	10	50	20	10

れる．

4　実験

4.1　実験概要

提案手法の性能を評価するために，性能比較実験を行った．比較手法として，タグの共起のみを用いる手法と類似画像のみを用いる手法を用いた．実験に用いたパラメータを表 1 に示す．

まず，実験 1 では，関連語抽出の性能に関する実験を行った．画像共有サイト flickr から検索質問 16 個を用いて収集した 5377 枚の画像，それらに付与されている 6805 個のタグを使用した．この画像とタグの集合からデータベースを構築して関連語抽出を行い，タグの共起のみを用いる手法から得られた G_{tag}，類似画像のみを用いる手法から得られた G_{img}，提案手法から得られた $G_{\cup}$，$G_{\cap}$ をそれぞれ評価した．評価尺度としては，結果として得られた正解タグの数 $|A|$，および適合率 $P = \frac{|A|}{|G|}$ を用いた．ここで，$|G|$ は処理結果として得られたタグの数である．また，正解の判定は著者が行った．

提案システムでは，ユーザが正解タグを選択しやすいように，一度の関連語抽出で，少数のタグを提示し，その中により多く正解タグを含めることを目的とする．従って，適合率，即ち抽出したタグにどれだけ正しいものが含まれているかが重要となり，再現率，即ちすべての正解タグをどれほど抽出できるかは，さほど重要ではない．

表 2　実験 1 で得られた正解タグの数 (個)

検索質問	G_{tag}	G_{img}	$G_{\cup}$	$G_{\cap}$
beach sunset	5	4	6	3
blue sky	4	4	4	4
cat house	1	2	2	2
city night	4	3	6	1
dance party	3	2	4	1
baby	1	0	1	0
hideki matsui	5	3	5	3
ichiro suzuki	8	6	8	6
pet dog	2	3	4	1
rock concert	4	4	4	4
sky clouds	2	2	2	2
summer beach	6	3	6	3
tokyo tower	4	3	4	3
tower bridge	7	5	7	5
winter snow	4	5	8	1
flower red	3	1	4	0
平均	3.9	3.1	4.7	2.4

実験 2 では，再検索の性能に関する実験を行った．実験 1 により得られた $G_{\cup}$ から人手で正解タグを選び，入力のタグに含め，再び関連語抽出を行った．これにより新たに得られたタグについて G_{tag}，G_{img}，$G_{\cup}$，$G_{\cap}$ をそれぞれ評価した．評価尺度としては，結果として追加された正解タグの数 $|B|$，および適合率 $P = \frac{|B|}{|G|}$ を用いた．その他の実験条件は，実験 1 と同様である．

4.2　実験結果と考察

まず，実験 1 の結果を表 2，3 に示す．表 2 に示すように，提案手法により得られた $G_{\cup}$ が，正解タグの数において最良の結果を得た．これより，タグの共起だけでは抽出できなかった関連語を，類似画像も考慮することで抽出が可能であることがわかった．特に検索質問「beach sunset」，「winter snow」では，色の分布が画像の内容をよく反映しているので，類似画像検索の精度が高くなった．このことにより，類似画像から新しい正解タグを多く抽出することができた．一方，検索質問「baby」，「cat house」では，類似画像から得られる正解タグの数は少なかった．色の分布だけでなく，新しい画像特徴量を吟味し類似画像検索の性能を向上させることができれば，このような検索質問でも，正解タグを多く抽出することができると考えられる．

また，表 3 に示すように，提案手法により得られた $G_{\cap}$ が，適合率において最良の結果を得た．これより，タグの共起と類似画像の両方から得られる関連語は，画像に付与される語である可能性が高いことがわかった．検索質問「ichiro suzuki」，「tower bridge」では，このような関連語が多く抽

表 3　実験 1 で得られた適合率

検索質問	G_{tag}	G_{img}	$G_{\cup}$	$G_{\cap}$
beach sunset	0.50	0.40	0.35	1.00
blue sky	0.40	0.40	0.27	0.80
cat house	0.10	0.20	0.13	0.40
city night	0.40	0.30	0.43	0.17
dance party	0.30	0.20	0.21	1.00
baby	0.10	0	0.06	0
hideki matsui	0.50	0.30	0.31	0.75
ichiro suzuki	0.80	0.60	0.62	0.88
pet dog	0.20	0.30	0.24	0.33
rock concert	0.40	0.40	0.40	0.40
sky clouds	0.20	0.20	0.17	0.25
summer beach	0.60	0.30	0.38	0.75
tokyo tower	0.40	0.30	0.27	0.60
tower bridge	0.70	0.50	0.50	0.83
winter snow	0.40	0.50	0.42	1.00
flower red	0.30	0.10	0.27	0
平均	0.39	0.31	0.31	0.57

表 4　実験 2 で得られた正解タグの数 (個)

検索質問	G_{tag}	G_{img}	$G_{\cup}$	$G_{\cap}$
beach sunset	2	3	4	1
blue sky	2	1	3	0
cat house	2	1	3	0
city night	4	3	6	2
dance party	2	1	2	1
baby	0	0	0	0
hideki matsui	6	4	6	3
ichiro suzuki	5	3	5	3
pet dog	2	0	2	0
rock concert	1	1	2	0
sky clouds	0	1	1	0
summer beach	3	2	4	1
tokyo tower	3	0	3	1
tower bridge	2	4	4	2
winter snow	0	1	1	0
flower red	1	0	1	0
平均	2.2	1.6	2.9	0.9

出され，またその中に正解タグを多く含んでいた．従って $G_{\cap}$ を用いることで，ユーザに関連度の高いタグを提示できると考えられる

次に，実験 2 の結果を表 4，5 に示す．表 4 に示すように，関連語の再検索を行うことで，タグの共起，類似画像それぞれから正解タグを追加できることがわかった．検索質問「hideki matsui」では，実験 1 よりも多くの正解タグを抽出

表 5　実験 2 で得られた適合率

検索質問	G_{tag}	G_{img}	$G_{\cup}$	$G_{\cap}$
beach sunset	0.20	0.30	0.22	0.50
blue sky	0.20	0.10	0.18	0
cat house	0.20	0.10	0.15	0
city night	0.40	0.30	0.38	0.50
dance party	0.20	0.10	0.11	0.50
baby	0	0	0	0
hideki matsui	0.60	0.40	0.38	0.75
ichiro suzuki	0.50	0.30	0.33	0.60
pet dog	0.20	0	0.11	0
rock concert	0.10	0.10	0.17	0
sky clouds	0	0.10	0.10	0
summer beach	0.30	0.20	0.22	0.50
tokyo tower	0.30	0	0.16	1.00
tower bridge	0.20	0.40	0.29	0.33
winter snow	0	0.10	0.08	0
flower red	0.10	0	0.07	0
平均	0.22	0.16	0.18	0.29

することができた．このように，タグを追加して再検索を行うことで，少数のタグの入力では抽出できなかった関連語を抽出することができると考えられる．

また，実験 1 の結果と同様に，正解タグの数においては $G_{\cup}$，適合率においては $G_{\cap}$ が最良の結果を得た．よって，関連語の再検索においても，提案手法は，比較手法に比べて関連語抽出の性能が優れているといえる．

なお．実験 1 の結果と比べて実験 2 による正解タグの個数と適合率が共に低下しているのは．再検索の難しさに起因する．しかし．これは再検索が有効でないことを表すのではない．再検索によって初回の検索では得られなかった新しいタグが得られることは．タグの充実に大きく寄与するといえる．

5　おわりに

本稿では，タグの共起と類似画像を用いたタグ付け支援システムを提案した．提案システムの特徴は，タグの共起だけでなく類似画像の関連語も抽出することで，多くの関連語をユーザに提示できる点である．

実験の結果，タグの共起と類似画像を用いることで，それぞれ単独で用いる手法より，多く関連語を抽出することができた．また，それらの関連語を用いて再検索を行うことで，新しい関連語を追加することができた．

今後の課題としては，より多くの正しい関連語を抽出するために，タグの共起，類似画像それぞれが持つ抽出性能の向上などが挙げられる．

参考文献

[1] A. Mathes: “Folksonomies - cooperative classification and communication through shared metadata”, Computer Mediated Communication, UIC Technical Report (2004).

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

[3] V.Gudivada. and V.Raghavan.: “Content-based image retrieval-systems”, IEEE Comput., Vol.28, No.9, pp. 18–22 (1995).

[4] Y. Rui, T. S. Huang and S. Mehrotra: “Relevance feedback techniques in interactive content-based image retrieval”, Proc. of Storage and Retrieval for Image and Video Databases (SPIE), pp. 25–36 (1998).

[5] E.V.Munson, Y. Tsymbalenko：“To search for images on the web，lood at the text，then look at the images”, Proc. of First International Workshop on Web Document Analysis, pp. 39–42 (2001).

[6] 出原博, 藤本典幸, 竹野浩, 萩原兼一：“WWW 画像検索における画像周辺の HTML 構文構造を考慮した画像説明文の抽出手法”, 電子情報通信学会技術報告, DE2005-136 (2005).

[7] 相良直樹, 砂山渡, 谷内田正彦：“HTML テキストの重要文を用いた画像ラベリング手法”, 電子情報通信学会論文誌, Vol.J87-D-I, No.2, pp. 145–153 (2004).

[8] 竹内謹治, 黄瀬浩一：“類似画像とキーワードを利用した web 画像の説明文抽出”, 情処研報, NL–171, pp. 7–12 (2006).

[9] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas and R. A. Harshman: “Indexing by latent semantic analysis”, Journal of the American Society of Information Science, Vol.41, No.6, pp. 391–407 (1990).

[10] Y. Rubner, C. Tomasi and L. Guibas: “The earth mover’s distance as a metric for image retrieval”, International Journal of Computer Vision, Vol.40, No.2, pp. 99–121 (2000).