

複数人での写真閲覧行為に基づく自動タグ付加システム

三島木 一磨^{†1} 戸田 真志^{†2} 川嶋 稔夫^{†2}

近年様々なメディアが簡単に保存できるようになった。写真もそのひとつであり、人々はそれらをどんどん保存しているが、それらの多くは整理されないまま保存されており、このままではそれらの写真は価値を失ってしまう。これらの有効な再利用のために、我々は複数人で写真を閲覧しおしゃべりをするシーンに着目した。このようなシチュエーションを解析することで写真に関する情報を抽出し、付加することが可能でないかと考えた。本論文では閲覧者の行動モデルと写真の情報モデル及び「タグの伝搬」モデルを定義する。そしてこれらのモデルに基づき画像処理によって写真の状態を解析しタグを付加するシステムを提案する。さらにそれを用いて実験を行い、モデルの正当性を検証する。

Automatic Tagging System through Discussing Photos

KAZUMA MISHIMAGI,^{†1} MASASHI TODA^{†2}
and TOSHIO KAWASHIMA^{†2}

Many media forms can be stored easily at present. Photographs, for example, can be easily stored even though most of them have not been edited. This means they will gradually lose their value and become essentially unusable. To make better use of photographs, we tried to make use of information provided by viewers who had seen and commented on them. We felt that analyzing this information would enable us to make maximum use of photographic data. To do this, we defined a “tag propagation” model and relationships between photos. We also proposed a system that uses image processing to analyze viewers’ handling of photos and how the photos are relevant to each other. We then validated our model by using it.

^{†1} 公立はこだて未来大学大学院
Graduate School of Future University-Hakodate

^{†2} 公立はこだて未来大学
Future University-Hakodate

1. はじめに

近年、ストレージの大容量化・低価格化によりあらゆるメディアが削除されることなく蓄えられるようになった。写真のデータに関しても例外はなく、デジタルカメラの普及によって誰もが手軽に写真を撮影し、保存するという行為が行われている。例えば函館市にはデジタルアーカイブ計画がある。この計画では過去の歴史的価値のある資料をデジタル化して保存しており、ここには十万枚程度 of 古写真も含まれている。当然一般家庭における環境でも、数千枚のオーダーで写真が保存されていることは想像に難くない。

写真というものは、撮影され、整理され、再利用（閲覧）されるというサイクルを経るものだと考えられる。しかし現在、写真は手軽に撮影・保存を行うことができるようにはなったが整理はほとんどされておらず、結局後に閲覧されることもなく埋もれていってしまう。前述したデジタルアーカイブの写真を含め、後に再利用されるべきこれらの写真群は、結局存在しないも同様の扱いで眠っている状態であり、このままでは徐々に価値を失いゴミ同様の扱いとなってしまう。これらの有効な再利用のためには何らかの手段で整理を行う必要がある。

写真を整理されることは少なくなったが、それに比べ閲覧する機会は多い。写真の閲覧は整理に比べると楽であり、楽しい行為である。例えば結婚式の後や、久しぶりに同期と会ったとき、帰省したとき、現像されたままの写真を撮りだし、広げて複数人で閲覧する事がよくある。これらのシチュエーションにおいては写真に写っている事柄のみからは分からないような情報を含め、それら写真に関する情報が飛び交う場となる。そこでこのような写真に関してディスカッションを行う場を用意し、そのシチュエーションを記録・解析することで、閲覧者達が特に情報を付加する意志を持たずとも、写真に様々なタグを付加して整理することが可能ではないかと考えた。

本研究で問題となる事柄はまず、閲覧者のどのような行動が、どのような情報に結びつくのかという事であり、次にそれらをどのように検出するかである。

本研究ではまず閲覧者の行動をモデル化する。次にディスカッションの行われているテーブル上の映像に対して画像処理を行う事でそれらを検出し、その結果付加された情報の品質に関して実験的に検証を行う。

2. アプローチ

2.1 提案手法

前述のシチュエーションにおいては写真に関する多くの情報がやりとりされる。そこで写真をテーブル上に広げてディスカッションを行う場を提供し、この状況をそのまま記録・解析することで、その場で発生する情報を取得し、未整理の写真に情報を付加する。

具体的には写真が広げられたテーブル上の映像を記録し、ディスカッションの際の閲覧者の写真に対する行為を解析して自動的に写真に情報を付加するシステムを提案する。

本手法を用いることで、情報を付加したい写真群に関して潜在的に知識を持った人々に、写真に関するディスカッションを行ってもらっただけで、質の高い情報を付加できると考えている。例えば古い写真群であれば、高齢者の方々に写真に関する話をしてもらうことで非常に自然に写真に情報を付加していくことができる。このシステムによって、写真群に関して少なからず知識のある人々がそれらに関するディスカッションをするという非常に自然な行為から、暗示的な情報を写真に付加し編纂することが可能となる。

2.2 提案システムの概要

まず閲覧者がテーブルに大量の写真を広げディスカッションを行う。その間、閲覧者の写真に対する行為をテーブル上方に設置したカメラで撮影し続ける。その映像から写真がどのように扱われたかをシステムがリアルタイムに解析し、写真に情報を付加していく(図1)。

付加する情報としては「タグ」の概念を用いる。「タグ」とはオンライン写真共有サービスやソーシャルブックマーキング等で良く用いられる用語で、写真、ブックマークに付加された情報の事を指す。

予め、閲覧する写真群の中にはタグが付加された写真が含まれている事を前提とする。その上で、閲覧者の写真に対するハンドリングから写真の関連を検出し、関連した写真同士の間で、後述するタグの伝搬を行うことでタグを付加していく。

本論文ではまず写真の関連性のモデル及び、タグの伝搬モデルを定義する。そしてこのモデルに基づき、関連性を出してタグを伝搬させるシステムを実装する。そしてこのシステムを用いた実験の解析結果について考察する。

3. 関連研究

情報編纂とテーブルトップシステムに関して多くの研究が行われている。

自動クラスタリングの研究として、Hilligesらは類似画像のクラスタを生成し、最良の画

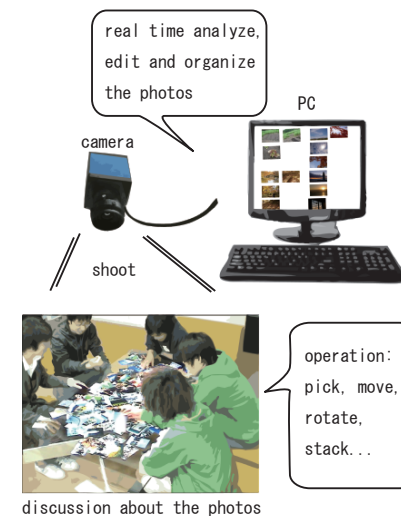


図1 提案システムの概要

Fig. 1 Overview image of the system

像を選出するシステムを作成している³⁾。また PhotoTOC⁵⁾の研究ではタイムスタンプとカラーヒストグラムから代表画像を選出するシステムを提案している。

これらの研究では画像の特徴によるクラスタリングとタイムスタンプの様な情報を用いた編纂が行われているが、情報の付加によるクラスタリングや編纂については議論されていない。

テーブルトップシステムは1993年に DigitalDesk⁷⁾で提唱された。DigitalDeskにおけるテーブルトップシステムは実世界と電子世界の橋渡しをするシステムであった。Koikeらはロータリーテーブルの概念を用いてテーブルトップシステムの問題であった情報の向きの問題を解決する試みを行った²⁾。シーケンシャルな配置や渦巻き状の配置に関して言及している。また Microsoft社はマイクロソフト・サーフェイスと呼ばれる商用サーフェイスコンピュータを発表しており、そのシステムはそのテーブル状の形状を活かして上に乗ったオブジェクトとのインタラクションを行うことも出来るシステムである。

デスク上のオブジェクトを継続的にトラッキングする研究として J.Kimらの研究がある⁴⁾。これはある時刻にある資料がどの位置に存在するかを全て記録することにより、コンピュー

タ内のデジタルな資料とデスク上のフィジカルな資料を結びつけ、物理的位置まで検索が可能にしたものであるが、あくまでデジタル・フィジカル間のコネクティビティを確立することに留まっている。オブジェクトの位置情報から新たな情報をオブジェクトに付加するような研究は行われていない。

多数の情報に複数人がアクセスすることで情報を付加し、整理する方法としては flickr⁸⁾ のような写真共有サービスが真っ先に思い浮かべられる。flickr では多くの人々が写真を閲覧し、タグやコメント、アノテーションをつけ、検索可能な巨大アーカイブを形成している。さらにそれらのタグやコメントに基づき検索を行ったり、似た嗜好を持つユーザや写真を知ることでもできる。しかしこれらは各個人がタグを付けた上でアップロードされているものであり、編纂の意志が最初から存在している写真群である。未整理のまま放置され、いつ、誰が、どこで撮ったのかわからなくなってしまったような写真に対しては有効にはたらない。またそのような大量の写真に対して能動的に情報を付加していくのはユーザに対して負荷がかかる作業であり、情報を付加しているという意識なくして情報を付加できる仕組みが必要とされる。

また、実世界とデジタル世界のメディアの取り扱いに関する研究が行われており、実世界のメタファとその入力方法はデジタル世界でのそれとは全く違うものであると結論づけられている⁶⁾。

4. モデル

写真に情報を付加するためにいくつかのモデルを提案する。写真の持つ情報モデル、写真の関連性モデル、閲覧者の行動と写真の状態のモデル、タグの伝搬モデルである。

4.1 写真の持つ情報とタグ

写真には、写真を画像として解析することで分かる情報の他、背景となる情報など様々な付加情報を持つ。flickr などの写真共有サービスでは、写真にタグという形でこれらの情報を付加することができる。この場合、写真が本来潜在的に持っている情報を人が明示的にタグという形で付加していると考えられる。

本研究でも同様に、写真にタグという形で情報を付加する事を考える。特に暗示的に、閲覧者が意識なくしてタグを付加できる仕組みとして、写真同士の関連性を定義し、関連している写真同士でタグを伝搬させるというアプローチをとる。

4.2 閲覧者の行動と写真の状態

閲覧者が閲覧中に行う行動と写真の状態について述べる。ここで、本研究における閲覧者

とは閲覧を行う写真群に対して断片的に、若しくは潜在的に知識を有する者を指し、写真に関して全く知らない人々が閲覧を行うような状況は想定していない。

沢山の写真が閲覧されている状況では、それらの写真に関する様々な情報が飛び交うと考えられ、それらの一部は写真の移動と閲覧者の行動から推測することができる。例えば誰かがある写真に注目したときには、大抵はその写真を近くに引き寄せる。もしそれほど気にならない写真であればその写真はそこに留まる。また、誰かがある写真に強い関心やコメントがある時に他の人にそれを伝えることがある。もしそれを聞いた人が関心を持てば、写真はその人の前にも移動する。写真への指差しも閲覧者の注目にに関するサインと考えられる。もし複数人が同じ写真を同時に指している事があれば、その写真がその時点での会話の主題となっていると考えられる。閲覧者が綺麗に写真を積み上げたり並べたりしていればグルーピングや比較を行っていると考えられる。ひとりがある写真を見ている時に、他の閲覧者がそれを見て関連する写真を引き寄せるといった行為も見られる。

閲覧者が写真を閲覧している際に写真がとる状態から、シグナルレベルで具体的には図 2-5 の様な情報が得られる。



図 2 位置 (移動)
Fig.2 Position(Moving)



図 3 重なり
Fig.3 Overlap



図 4 回転 (角度)
Fig.4 Rotation(angle)



図 5 指されている
Fig.5 Pointed

我々は今回、前述した閲覧者の行動モデルのうち、関連する写真を積んだり並べたり、同時に閲覧するという行為に着目し、近くに置かれた写真同士には関連性があると定義した。

4.3 写真同士の関連の定義

写真同士の関連性を次のように定義する。

ある写真 P_1 の持つ情報の集合を I_1 , 写真 P_2 の持つ情報の集合を I_2 としたとき,

$$I_1 \cap I_2 \neq \phi \quad (1)$$

のとき, P_1 と P_2 は関連しているという.

4.4 写真の関連とタグの伝搬

写真同士が関連しているとき, 前述の定義より, $I_1 \cap I_2 \neq \phi$ である. 写真 P_1 と P_2 が, 情報 I を介して関連しているとき, P_1 が I を具体的なタグ T として保持していたとすると, この T を P_2 に付加することができる. そこで, 関連した写真同士は何らかの共通のタグを持つと仮定する.

このように, ある情報を介して関連している 2 つ以上の写真同士のうち, そのタグを持っている写真から持っていない写真にそのタグを付加することを「タグの伝搬」と表現する. 提案システムではこの「タグの伝搬」を用いて, タグのない写真にタグを付加する.

タグの伝搬の方式に関して, 表 1 にまとめた. 一番左のカラムから, 実際に付加されているタグと写真の関係, その時に理想的な伝搬の仕方の図, 分類名となっている.

今回はこの方式のうち, 全ての伝搬は相互全伝搬であると仮定した. 今後, 細かな伝搬方式の推定等も行う必要があると考えているが, 本論文では取り扱わない. この定義では閲覧を通してかなり大量のタグがそれぞれの写真に付加され, しかもどの写真も同じようなタグを持つようになると考えられる. またそのタグ群の中には当然, 実際には正しくないタグも含まれる事が予想される. しかし何度も様々な写真と共に閲覧を繰り返し, タグに重み付けを行っていくことで, 付加された情報の精度が向上していくと考えられる.

重みとして現在 3 種類の重みを設定している. 空間重み, 時間重み, 伝搬数重みである. 空間重みは, ある写真から別の写真までの物理的距離に基づいて計算される. 時間重みは時間経過に基づいて計算される. 伝搬数重みはタグの伝搬の回数に基づいて計算される. これら 3 つの重みが乗算され, 最終的な重みとなってタグに付加される. 伝搬先に既にそのタグが存在する場合, 重みは加算されていく.

5. システム

5.1 システム要件

システム概要で述べた状況を解析するためには

- テーブル上の写真の特定
- テーブル上の写真の位置姿勢の判定
- テーブル上の写真のトラッキング

- 関連性の検出とタグの伝搬

といった機能を実現する必要がある.

我々は解析すべき事柄として, 最終的なテーブル上の状態のみには大きな意味は無いと考えている. 自然な写真の閲覧では会話の流れに沿ってテーブル上の状態は刻々と変化していくため, テーブル上の写真の状態を作業の結果の最終的な状態のスクリーンショットとしてではなく, その時々における状態の流れとして捉えるべきだと考えている. ここに映像を逐次的に解析し写真をトラッキングする必要性が生まれる.

5.2 システムの設計

図 6 は処理の流れを表したものである.

5.2.1 写真への枠線と余白の付加

まず写真には検出を簡単にするために枠線と余白を付加した. 枠線の太さは経験的に, 写真の短辺の 2.5% の太さにした. 余白は短辺の 15% の幅に設定した. これらの写真を印刷し, 閲覧に用いる.

5.2.2 写真の検出

テーブル上方にカメラを取り付け, 閲覧を行う. まず, システムはカメラからテーブル上の映像を得る. 現在は予め録画した映像に対して解析を行っている. この取得は 15 フレーム (0.5 秒) に 1 度行う. 得られた映像を二値化し, 膨張・収縮をかける. そして輪郭検出を行い輪郭を含む最小の矩形を検出する. 検出された矩形は大きさや形によりフィルタリングされ, 写真として検出される (図 7). この時点で写真の位置姿勢が明らかになる.

5.2.3 手の領域の検出

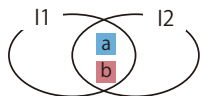

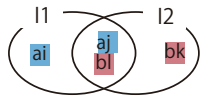

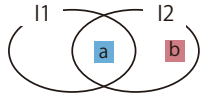

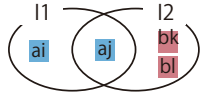
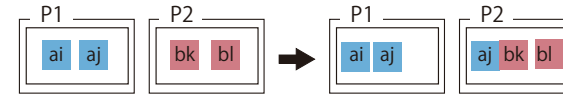
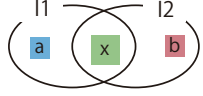
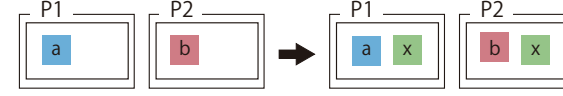
写真が閲覧者の手に持たれているか, テーブル上に置かれている状態であるかを判断するために, 簡易に手の領域の検出を行っている. まずカメラにより取得されたフレームの色空間を HSV 空間に変換し, Hue の値が 0-30 の部分を肌色領域とする. この値は経験的に決定しているもので, 環境により適宜変更している. 次にこの肌色領域から十分な面積 (今回は 5000px) をもった領域を手の領域とする. 手の領域の重心から 80px 以内のエリアに写真が入っている場合は手に持たれている可能性があるとして判断し, 伝搬は行わない.

5.2.4 SURF による写真の同定

SURF¹⁾ は画像中の点の周囲の窓内部のエッジ方向ヒストグラムに基づく特徴記述子である. 抽出される 64 次元の特徴は二次元における回転・平行移動・スケールに不変であり, 輝度・コントラスト・部分的隠れに頑健である. 今回写真の同定にこれを用いている.

システムは写真の ID と対応する SURF 特徴量の参照テーブルを持つ. これは予めシス

表 1 タグ伝搬の方式
Table 1 Patterns of tag propagation

分類名	写真同士の関連の仕方	理想的なタグの伝搬
1: 相互全伝搬	 $\forall a, \forall b \in X$	
2: 相互選択伝搬	 $\text{Some of } A, B \in X$	
3: 有向全伝搬	 $\forall a \in X$ $\forall b \notin X$	
4: 有向選択伝搬	 $\text{Some of } A \in X$ $\forall b \notin X$	
5: 伝搬なし (生成, 推定)	 $\forall a, \forall b \notin X$	

テム内の写真から SURF 特徴量を抽出し, ID と共に保存しておいたものである。ここでテーブル上で検出された写真の領域から SURF 特徴量を抽出し, 参照テーブル内のデータを比較し写真を同定する。図 8 はテーブル上のある写真と PC 内のデータとを対応づけた例である。

写真領域が検出されると, その領域の特徴点とデータベース上の全ての画像と対応点を取り, もっとも対応点の多かったデータベース上の画像とテーブル上で検出された画像が同じ

写真であると判断する。

5.2.5 写真のトラッキング

システムは現在の状態テーブルという別の表を持つ。これは写真の ID, 位置, 角度と SURF 特徴量を要素として持つ。写真が検出され同定が済んだ場合, 現在の状態を更新する。この際に写真が検出された位置と, 最新の状態テーブルにおける写真の位置を比較し, 移動や角度の変化を検出する。今回は 30px 以上の変化があった場合写真の移動が開始した

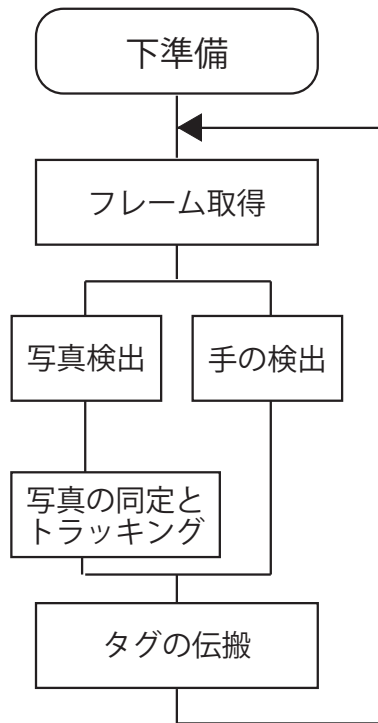


図 6 処理の流れ
Fig. 6 Process overview

と判断し、次のフレームで再度移動が無ければ移動が終わったと判断している。

5.2.6 関連性の検出とタグの伝搬

写真の移動が検出された際に、その写真が何らかのタグを持っていた場合、それらの位置関係や時間経過に基づいて重みが計算され、タグの伝搬が行われる。今回の重みの定義は表 2 に示した通りである。

これらの重みは乗算され、その後タグに付加される。伝搬数重みに関して、伝搬回数が 1 回未満なら 1 とは、すなわち、元々付加されていたタグであれば伝搬するということである。つまり今回は、閲覧開始時から元々タグが付加されていた写真が、移動直後に半径 80px 以内に、その場に置かれてから 60 秒以内の写真があった場合にのみ、重み 1 のタグが伝搬す

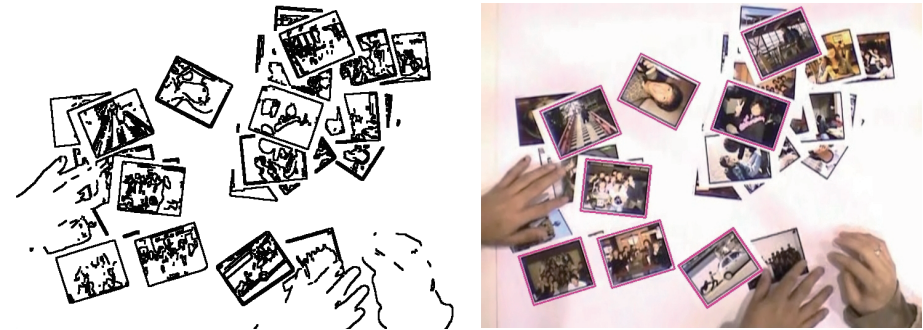


図 7 写真検出
Fig. 7 Detection of photos

表 2 重みの定義
Table 2 Definitions of weights

重み種別	関数
空間重み	ある写真が半径 80px 以内にあるなら 1, それ以外は 0
時間重み	ある写真の最後の移動完了からの経過時間が 60 秒以内なら 1, それ以外は 0
伝搬数重み	伝搬回数が 1 回未満なら 1, それ以外は 0

ることになる。

5.3 ハードウェア

テーブル上の状態を撮影するカメラとして一般的なビデオカメラを用いた。これは 720x480 ピクセルの映像を 30fps の速度で撮影できるものである。カメラは三脚と自作のアームを用いてテーブル上 120cm の位置に設置した。テーブルは表面が白色で拡散反射性である。演算に用いたコンピュータは一般的なノート PC であり Core2 Duo 2.2 GHz の CPU と 2 GB の RAM を搭載するものである。

システムとして写真の挙動を検出するために巨大なタッチディスプレイを使う方法も考えられる。しかしそれは非常に高価であるし、数百枚の写真を十分な解像度で扱えるかということにも疑問が残る。さらにタッチディスプレイでは埋もれた写真を直感的に扱うことは難しい。

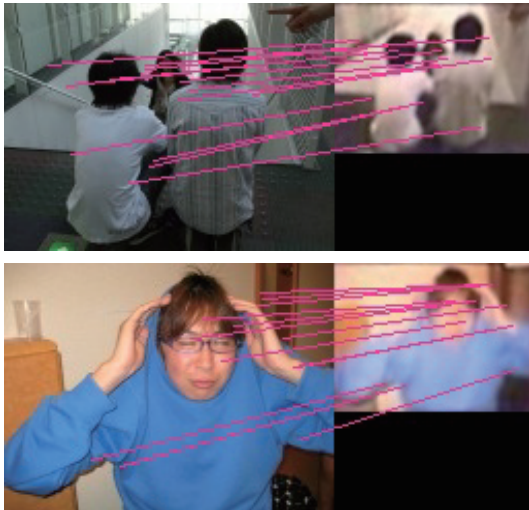


図 8 テーブル上の写真と PC 内のデータの対応付け
Fig.8 Photo on the table matched the data stored in PC

6. 実験

我々のモデルによって正しい関連付けが得られるかを調べるために、実験を行った。

6.1 実験設計

実験の概要は以下である。

- 写真：大学生 A さんの写真 40 枚。
- 被験者：A さんの友人の大学生 2 名。互いに仲が良い。
- 時間：15 分間
- 指示：「15 分程度、会話をしながら自由に閲覧して下さい。写真について知っていることがあれば共有してください」

実験に用いた写真は大学生 A さんの撮影した写真群であり、被験者は A さんの友人 2 名である。これは閲覧者は閲覧する写真に関して少なからず知識を持つという前提を満たす。被験者同士は互いに仲が良い、これは閲覧中の会話を活発に行ってもらいたい意図がある。閲覧時間は経験的に 15 分間としている。

指示としては、会話をしながら自由に閲覧をするように述べ、また各人の知識を閲覧に

反映してもらうために互いの知識を共有するように述べた。特に整理や情報の付加を行うような指示は行っていない。

6.2 実験の解析

前述の実験の映像をシステムを用いて解析する。まず、実験に用いた全ての写真に対して何らかのタグが付加されている状態で解析を行う。それらを重み 1 のタグとみなして実験を行い、解析結果に対して各タグにより検索をかける。我々のモデルが閲覧行動に現れていれば、予め付加されていたタグの重みが高まり、検索結果の上位に正しい写真が現れるはずである。

次に、タグの数が元々付加されていた 30%程度の量になるようにいくつかのタグを削除し、「いくつかの写真にはタグが付加されているが、他には何の情報もない」という、本研究で想定している状態で再度解析を行い、正しくタグが付加されているかを調べる。

6.3 実験結果と評価

図 9 は実験を解析した後にいくつかのタグを用いて検索をかけた結果、検索された上位 5 枚の写真を表している。1 行目は検索に使ったタグ、2 行目は解析前にタグを削減しているかいないか、3 行目は検索の上位 5 枚の写真である。検索の際のランキングはタグの重みによって決定されており、つまり重みが大いほど上位にランクされる。青い背景で示された写真が、全ての写真にタグが付加された状態のときに検索に用いたタグが解析前から予め付加されていた写真、つまり正解の写真である。

まず、全てのタグが既知の場合の結果に関して、図に示した通り、解析前の状態でそのタグが付加されていた写真がランキングの上位に入っている。これは我々のモデルの通り、閲覧者が関連する写真同士を近くに置き、タグが伝搬したことを示していると考えられる。次に、タグの総量が全体の 30%程度になるように削減した後の結果についても同様に、既知であったタグと同じタグが付加されている。これらの結果から、「閲覧者は関連した写真同士を近くに置く」という行動が特に整理の指示のない閲覧中に発生しており、それに基づいたタグの付加が可能であると考えられる。

また、タグを削減した後の方が検索結果として正しい結果が多く得られている場合もある。これは、全ての写真にタグが付加されていた場合、ひとつの写真にいくつかの種類タグが付加されていることが良くあり、そのとき今回の仮定である全伝搬によって関係の無いタグも同時に伝搬することで、ノイズが増えてしまったのではないかと考えられる。これは表に示した検索結果に同じ写真が何度か登場することにもつながっていると思われる。今後、伝搬するタグの推定などを行うことで精度の向上が期待できる。

検索に用いたタグ	カラオケ		プロジェクト		誕生日	
解析前の状態	全てのタグが既知のとき	いくつかのタグを削減したとき	全てのタグが既知のとき	いくつかのタグを削減したとき	全てのタグが既知のとき	いくつかのタグを削減したとき
検索された結果上位5枚の写真						

解析前に検索に用いたタグが付与されていた写真群

図9 検索結果の一例

Fig.9 Example of retrieving results

7. おわりに

暗黙的に写真にタグ付けを行うためのアプローチとして、写真を用いたディスカッションを画像処理によって解析し、関連性を定義する手法を提案した。

閲覧者の行動と写真の関連性をモデル化し、我々のシステムを用いて写真を用いたディスカッションを解析し、写真にタグを付加した。

結果として既知のタグと同様にいくつかの写真に対してタグを付加することに成功し、「いくつかの写真にタグが付加されており、その他には情報がない」という状態において、「閲覧者は関連した写真同士を近くに置く」という行動に基づいたタグの付加は可能であると考えられる。

ノイズの削減のためにタグの伝搬モデルに関して洗練していく必要がある。閲覧者の行動と写真の関連性をさらに詳しくモデル化して伝搬モデルに反映したり、伝搬モデル、伝搬するタグを推定することで、さらに情報の精度が高まると考えている。

参考文献

- 1) Bay, H., Tuytelaars, T. and Gool, L. J.V.: SURF: Speeded Up Robust Features, *ECCV (1)* (Leonardis, A., Bischof, H. and Pinz, A., eds.), Lecture Notes in Computer Science, Vol.3951, Springer, pp.404–417 (2006).
- 2) HidekiKoike, ShintaroKajiwara, K.F. and Sato, Y.: Information Layout and Interaction on Virtual and Real Rotary Tables, *Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer System*, pp.95–102 (2007).
- 3) Hilliges, O., Kunath, P. and Pryakhin, A.: *Browsing and Sorting Digital Pictures Using Automatic Image Classification and Quality Analysis*, 3, pp.882–891, Springer-Verlag (2007).
- 4) Kim, J., Seitz, S.M. and Agrawala, M.: Video-based document tracking: unifying your physical and electronic desktops, *Proceedings of the 17th annual ACM symposium on User interface software and technology*, Santa Fe, NM, USA, ACM, pp. 99–107 (2004).
- 5) Platt, J.C., Czerwinski, M. and Field, B.A.: PhotoTOC: Automatic Clustering for Browsing Personal Photographs, Technical report, Microsoft Research (2002).
- 6) Terrenghi, L., Terrenghi, L., Kirk, D., Sellen, A. and Izadi, S.: Affordances for manipulation of physical versus digital media on interactive surfaces, *Proceedings of the SIGCHI conference on Human factors in computing systems*, San Jose, California, USA, ACM, pp.1157–1166 (2007).
- 7) Wellner, P.: Interacting with paper on the DigitalDesk, *Communications of the ACM*, Vol.35, No.7, pp.87–96 (1993).
- 8) Yahoo!Inc.: Flickr, <http://www.flickr.com/>.