

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 6 日現在

機関番号：14401

研究種目：若手研究(B)

研究期間：2011～2013

課題番号：23700117

研究課題名(和文) 適合性フィードバックによる人物の共起関係を考慮した画像検索システムの構築

研究課題名(英文) Development of Image Retrieval System by Considering People Co-occurrence Relations through Relevance Feedback

研究代表者

新田 直子(Nitta, Naoko)

大阪大学・工学(系)研究科(研究院)・講師

研究者番号：00379132

交付決定額(研究期間全体)：(直接経費) 3,300,000円、(間接経費) 990,000円

研究成果の概要(和文)：家族や友人などが頻繁に共に撮影される集合写真コレクションから特定人物を含む画像を人物の共起関係を考慮して検索するシステムを構築した。画像内の各人物の顔の見た目とその共起関係を同時に表現する人物共起特徴空間において、検索対象人物を含む画像と含まない画像を分類するクラス分類器を訓練する。このクラス分類器により検索対象人物を含むと判定された画像を検索結果としてユーザに提示し、ユーザが少数の画像にフィードバック情報として与える正解/不正解情報に基づきクラス分類器を再度訓練する。1791枚の集合写真からの32名の人物の画像検索の結果、5回のフィードバック後、平均精度が20%から70%に向上した。

研究成果の概要(英文)：Group photos often contain socially related people such as family and friends. We developed a system for retrieving images containing a specific target person from a group photo collection by considering people co-occurrence relations. A Bag of People (BoP) feature, which represents both the facial appearances of persons and their co-occurrence relations, is extracted from each image. By using the BoP features, a classifier for classifying images into two classes, images containing the target person and other images, can be trained. The images determined to contain the target person by the classifier are presented to the user as the retrieval results. The classifier is re-trained by a small number of correctly retrieved images which are selected by the user. When retrieving images of 32 persons from 1791 group photos, the mean average precision improved from 20% to 70% after five feedback iterations.

研究分野：総合領域

科研費の分科・細目：情報学/メディア情報学・データベース

キーワード：人物画像検索 適合性フィードバック 共起関係 クラス分類 能動学習 構造化

1. 研究開始当初の背景

個人が日常的に撮影する画像では人物が被写体となることが多く、大量の画像が蓄積されるにつれ、特定人物の画像を高精度に検索する技術が重要となる。人物画像検索は一般に、画像中の人物が誰であるか同定する顔認識技術を用い、クエリ画像として与えられる検索したい人物の顔画像から、顔の部位(目、鼻、口等)の相対位置や大きさ、形などを表す特徴量を抽出し、見た目の類似度が高い顔領域を含む画像を選別することにより実現される。ところがこれらの特徴量は、人物の顔の向き、表情や一部の隠れ、撮影環境の照明条件など、さまざまな要因により変化する。そこで、顔の見た目以外に、画像の撮影日時や位置情報、人物の服装や人間関係などのコンテキスト情報を用いて認識精度を向上させる試みがなされている。特に人間関係は比較的長期に渡り変化しない情報であるため、長期間に渡り蓄積される画像集合からの検索において有用と考えられる。

出現する各人物に対する名前がラベルとして付与された集合写真が大量に存在する場合、各人物の顔領域から抽出された特徴量と人物名との関係、及び、人物間の同一画像における共起頻度を学習できる。しかし、ラベルを予め与えるユーザの手間などを考慮すると、高品質なラベル付き画像を大量に準備することは簡単ではない。一方、画像検索精度を改善させる代表的な手法に、ユーザが検索された少量の画像に対して与える正解/不正解の情報を用いて特徴量の重みや分類器を学習する適合性フィードバックがある。そこで本研究では、ラベルなしの大量の集合写真を対象とし、少数の画像に対してユーザが判定した、特定人物を含む/含まないという情報をラベルとして用いる適合性フィードバックにより人物の共起関係の学習を試みる。

2. 研究の目的

適合性フィードバックにより人物の共起関係を学習する場合、以下のような問題が考えられる。

(1) クエリ画像や適合画像においてクエリ人物と共に写る人物は、他の画像ではクエリ人物と共に写っていない、他の画像において他の人物とより頻繁に共に写っている、といった可能性があり、必ずしも検索に有効な共起人物とはならない。このようなクエリ画像以外の画像に関する情報は、検索開始時には不明である。

(2) 適合性フィードバックにおけるユーザの操作は、検索結果として提示された各画像からの適合画像の選択が一般的である。この場合、適合画像中のどの顔領域がクエリ人物か、また周囲の人物が誰か、については正しい情報が得られない。

(3) ユーザは検索結果中のすべての適合画像を選択する、もしくは正しい適合画像のみ

を選択するとは限らない。

以上を踏まえ、本研究では、適合性フィードバックによる人物の共起関係の逐次的学習方法、及び学習した人物の共起関係を用いた画像検索方法の検討を主な研究課題とする。また、インタフェースを用いた検索精度の検証も行う。

3. 研究の方法

本研究では、ユーザが検索したい特定人物を検索対象人物、大量の集合写真を画像コレクションと呼ぶ。本研究で対象とする画像検索は、ユーザにより検索対象人物を含む画像がクエリとして与えられたとき、画像コレクションから検索対象人物を含む画像を検索することが目的であり、画像中の各人物が誰であるか同定する必要はない。そこで、人物同定を行わず、画像内のすべての人物の顔の見た目と顔の見た目の共起関係を表現する人物共起特徴量を提案する。

まず、画像コレクション中の画像に含まれる代表的な顔の集合を得るために、これらの画像に含まれるすべての顔領域を顔の見た目の類似度に基づいてクラスタリングする。各クラスタの中心となる顔の集合を顔コードブックと呼び、与えられた画像コレクションの顔の辞書とする。各画像の人物共起特徴量は、辞書中の各顔コードに対し、画像中のすべての顔領域から抽出された顔特徴量との類似度の最大値を算出することにより得られる。つまり人物共起特徴量は、顔の辞書を用いて、どの代表的な顔の組が画像内に共起しているかを表現しており、画像内の人物数によらず任意の画像に対し、顔コード数の次元の特徴量となる。

初回の検索時においては、クエリ画像以外の画像に関する情報がないため、クエリと画像コレクション中の各画像の人物共起特徴量のコサイン類似度を算出し、検索結果としてコサイン類似度の大きい順に上位の画像をユーザに提示する。ユーザは提示された画像の中から、検索対象人物が含まれた画像を適合画像として選択する。また、提示された画像のうち、選択されなかった画像を非適合画像と呼ぶ。

これらの適合画像/非適合画像を正例/負例とし、人物共起特徴空間において検索対象人物を含む画像/含まない画像を分類する2クラス SVM(Support Vector Machine)分類器を訓練する。以降の検索においては、分類器により検索対象人物を含むと判定された画像を検索結果としてユーザに提示する。検索結果に対するユーザからのフィードバック、フィードバック結果に基づいた分類器の訓練の繰り返しにより、検索対象人物の画像検索に有効な顔の見た目と顔の見た目の共起関係が逐次的に学習され、画像検索精度の向上が期待される。

ここで、分類器の訓練に利用できるのは、ユーザからフィードバックが与えられた少

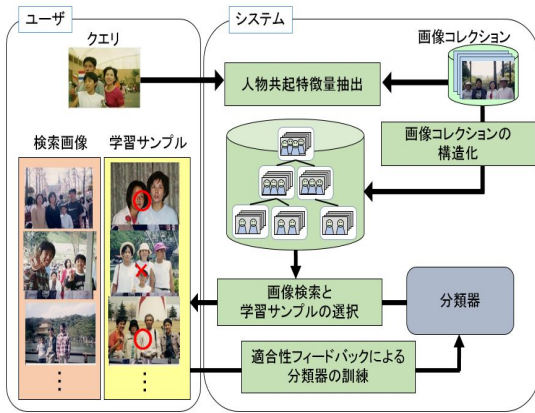


図 1: 提案手法の全体図

数の画像のみである。そこでさらに効果的な学習を行うため、フィードバック毎にクラス分類器の訓練に最も有効な学習サンプルを選択する能動学習を併用し、検索結果とは別に、能動学習により選択された画像の集合を学習サンプルとしてユーザに提示する。ユーザは、検索結果ではなく、学習サンプルから適合画像を選択することにより、効率的に検索対象人物の画像を得ることができる。

一方、提案手法により画像検索を行う際、画像コレクション内のすべての画像に対し、分類器により検索対象人物を含むか否か判定する。この場合、画像コレクション内の画像枚数が多くなるに伴い計算量が増加する問題が生じる。ここで画像検索の効率化の方法の一つに、画像データベースをあらかじめ意味内容に基づき階層構造化し、最初に上位階層において粗く探索し、画像の候補を絞った後に下位階層において詳細に検索する方法がある。本研究で対象とする集合写真には家族や友人などの高い頻度で共起する特定の人物グループが存在し、人物グループは、家族を上位階層としたとき、その部分集合である両親や子供などが下位階層となるような階層構造を持つ。よって、あらかじめ画像コレクションを人物グループに基づき階層構造化し、クエリ画像以外に画像コレクション内の画像に関する情報も考慮した効率的かつ高精度な検索を実現する。

具体的には、まず、画像間の人物共起特徴量の類似度に基づくクラスタリングにより、同一人物グループが撮影された画像集合を抽出する。得られた各画像集合の特徴量を画像集合内の画像の人物共起特徴量から算出し、画像集合間の特徴量の類似度に基づき、家族と子供のように一方が他方の部分集合である人物グループが撮影された画像集合同士をリンク付ける。画像検索時には、上位階層から、画像集合の特徴量を用い、クエリとの比較もしくは分類器により検索対象人物を含む画像集合が否かを判定する。検索対象人物を含むと判定された画像集合については、さらに下位階層の画像集合の判定を行う。最後に検索対象人物を含むと判定された画像集合中のすべての画像を検索対象画像

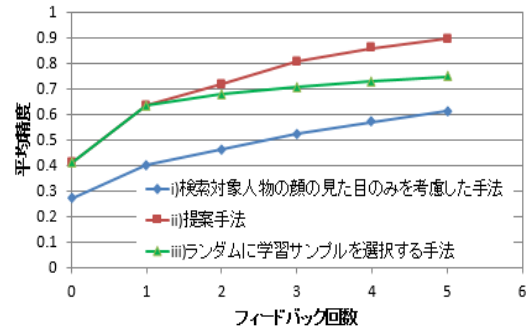


図 2: 人物共起特徴量と能動学習の評価結果

とし、最終的な判定を行う。
提案手法の全体図を図 1 に示す。

4. 研究成果

まず、顔の向きや大きさ、表情、経年変化や一部の隠れ、撮影環境の照明条件などの影響で見た目による認識が困難な顔を含み、かつ多様な人間関係を含むよう、計 140 名の人物を含む 1000 枚の画像を収集した。画像内のすべての顔領域及び顔特徴点を人手で指定し、顔の見た目を表す顔特徴量を抽出した。画像コレクション中のすべての顔領域の顔特徴量から 120 個の顔コードから成る顔コードブックを作成し、各画像に対し人物共起特徴量を抽出した。この中から家族や友人と頻繁に共起する 45 名の人物を検索対象人物として選び、これらの人物を含むのべ 1773 枚の画像をクエリとし、検索結果を正解画像がどれだけ上位に集中しているかを表す平均精度 (0 から 1 の値をとり、正解画像が上位に集中しているほど 1 に近い) により評価した。ただし、正解画像とは画像コレクション内の検索対象人物を含む画像のことである。また、学習サンプルとして提示する画像の枚数は 15 枚とし、ユーザによる誤判定の影響をなくすため、各フィードバックにおいて自動で正しい判定結果を与えた。

人物共起特徴空間で能動学習を用いて分類器を訓練する提案手法を、顔特徴空間で同様に分類器を訓練することにより検索対象人物のみを考慮した検索手法と比較した結果、図 2 に示すように、人物共起特徴量を用いた場合の方が高い平均精度が得られた。また、人物共起特徴空間において、ランダムに選択した学習サンプルを用いて分類器を訓練した場合の結果も図 2 に示す。学習サンプルの選択方法によらず、人物共起特徴空間において訓練した分類器を用いた方が、平均精度が高い。つまり、集合写真を対象とした特定人物を含む画像検索において、人物の共起関係をコンテキスト情報として利用することは非常に有効と言える。しかし、能動学習を用いた場合の方が、フィードバックを繰り返すごとに平均精度が著しく向上しており、能動学習により共起関係が効率的かつ高精度に学習されることが示された。

図 3 にクエリと正解画像例、表 1 に、図 3



(a) クエリ



(b) 正解画像

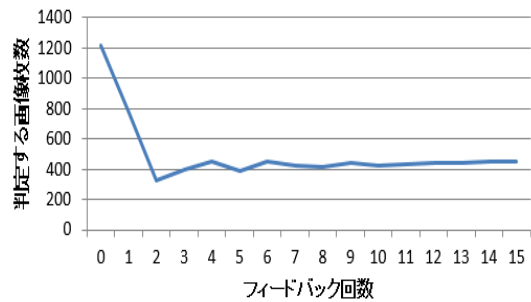
図 3: クエリと正解画像例

表 1: 検索結果例

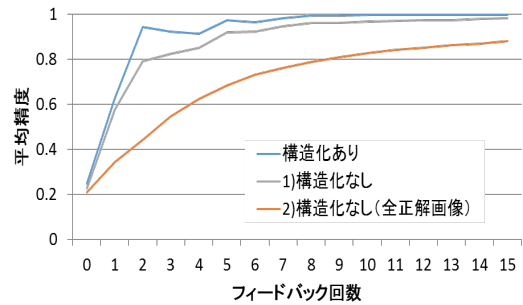
ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
i)の順位	1	2	3	4	5	6	7	8	9	10	13	46	66	103	128	133
ii)の順位	5	7	6	12	14	15	3	1	11	8	9	2	13	10	17	4

に示すクエリを用いて、i)検索対象人物のみを考慮した検索手法、ii)人物の共起関係を考慮した提案手法により、5回フィードバックを繰り返して検索を行った後の正解画像の順位を示す。ただし、赤文字は正しく検索された正解画像の順位、黒文字はそれ以外の正解画像の順位を表す。また、図3内の画像において、検索対象人物は赤枠で囲み、検索対象人物と二回以上同じ画像内に共起した人物については、同一の人物を同一の形状の緑枠で示す。表1に示すように、i)によってID1~8の正解画像が検索された。一方、ii)ではID15を除いたすべての正解画像が検索された。人物の共起関係の考慮により、ID13のように検索対象人物が子供の時の画像や、ID16のように検索対象人物の顔が正面を向いていない画像など、検索対象人物の顔の見た目が照明変化や隠れ、経年変化によって大きく変化した画像であっても検索できたと考えられる。

次に、190名の人物グループを構成する226名の人物を含む1791枚の画像を用いて画像コレクションの構造化の有効性を検証した。各人物グループの画像は1~50枚、121名の人物グループは他の人物グループの部分集合である。2個以上の人物グループに属する32名を検索対象人物として選び、各人物を含む計107枚の画像をクエリとして、画像コレクションを構造化する場合、しない場合の検索結果を、検索対象人物を含むか判定を行った画像の枚数及び平均精度により評価した。ただし、各画像内の顔領域及



(a) 判定した画像枚数



(b) 平均精度

図 4: 画像コレクション構造化の評価結果

び顔特徴点はFaceSDK(<http://www.luxand.com/facesdk/>)を用いて自動で抽出し、各画像の人物共起特徴量は、画像コレクション中のすべての顔領域の顔特徴量から作成した260個の顔コードから成る顔コードブックを用いて抽出した。

図4(a)に示すように、構造化した場合に判定を行った画像枚数は、初回時の1219枚から、フィードバック2回で約400枚に減少した。また、図4(b)に構造化する場合、しない場合の平均精度を示す。ただし、構造化しない場合は画像コレクション中の全正解画像の順位が算出できるが、構造化する場合は検索対象画像に対してしか順位を算出できない。よって、構造化しない場合の平均精度は、1)構造化する場合の検索対象画像内における正解画像と2)全正解画像に対して算出した。図4から、画像コレクションの人物グループに基づく構造化は、構造化しない場合に上位に検索された画像の順位を落とすことなく、計算時間を短縮したことが分かり、集合写真からの特定人物画像の検索において、人物の共起関係は検索効率・精度の両方を向上させる有用な情報であることが示された。

図5に、クエリにおいて検索対象人物の顔領域検出が失敗した場合の検索結果例を示す。クエリにおいて検索対象人物と共起する人物の顔領域のみが検出されたため、初回の検索においては、この人物が撮影された検索対象人物を含まない画像が上位に提示された。ユーザからのフィードバックにより検索対象人物や検索対象人物と共起する別の人物を学習することにより、クエリとは異なる様々な人物グループが撮影された正解画像

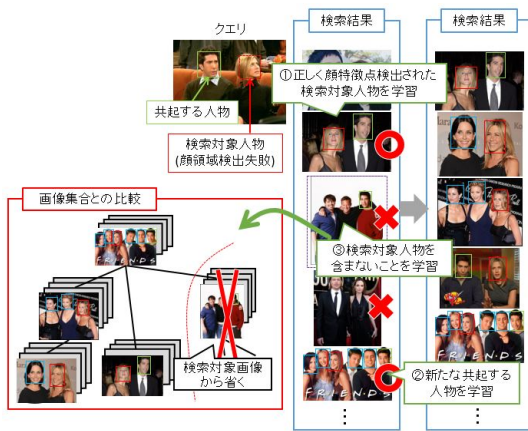


図 5: 構造化した画像コレクションからの検索結果例

を検索結果の上位に提示した。また、非適合画像から検索対象人物と共起しない人物を学習した後、検索対象人物を含まない人物グループの画像を画像集合に対する判定段階で一度にまとめて削除し、検索対象画像が減少した。

最後に、ユーザによるクエリ画像の入力、システムによる検索結果及び学習サンプルの提示、及びユーザによる適合画像の選択を行うためのインタフェースを構築し、適合性フィードバックにおけるユーザの誤判定による影響を検証した。被験者 4 名がそれぞれ、家族や友人と頻繁に出現する 10 名の検索対象人物を含む画像 10 枚をクエリとして 1000 枚の集合写真から検索を行った結果を、学習サンプルに対して正しく適合画像が選択された際の検索結果と比較した。図 6、7 にそれぞれ、フィードバック毎に提示される学習サンプルに対する被験者の判定の平均正答率、フィードバックごとの検索結果の平均精度を示す。各フィードバックでおよそ 85% から 98% の正答率が得られた。つまり、被験者は 15 枚の学習サンプルとして提示された画像のうち、1、2 枚程度の画像に対して誤判定したことを示す。特に、人物の顔の表情が大きく変化した画像や人物の年齢がクエリとは異なる画像に対する誤判定が多く見られた。しかし図 7 より、この程度の誤判定は検索精度に大きく影響しないことが分かる。この理由として、能動学習により分類器の決定境界に近い画像が学習サンプルとして選択されているため、ユーザが少数の画像に対して誤判定しても、分類器の決定境界は大きく変化しないためと推定される。このように、ユーザによる誤判定に対する提案手法のある程度の頑健性が示された。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表](計 5 件)

- [1] 中井優次郎, 新田直子, 馬場口登, “集合写真データベースの構造化に基づく画像

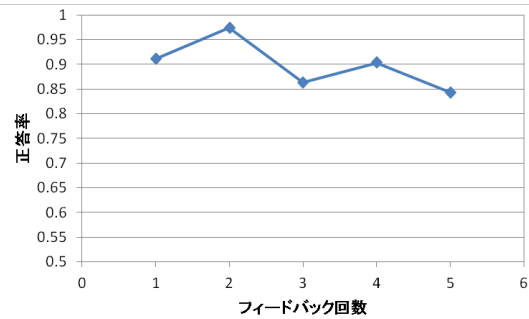


図 6: ユーザのフィードバックの正答率

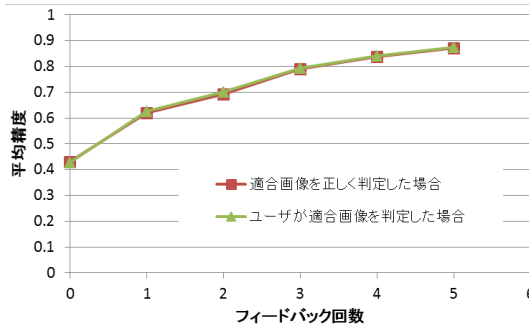


図 7: ユーザの誤判定の影響評価結果

検索”, 電子情報通信学会技術研究報告, PRMU2013-200, pp. 191-196, March 2014.

- [2] 志水和也, 中井優次郎, 新田直子, 馬場口登, “人物共起特徴量を用いたクラス分類に基づく集合写真の検索”, 画像の認識理解シンポジウム (MIRU2012), OS11-03, August 2012 (査読有).
- [3] K. Shimizu, Y. Nakai, N. Nitta, and N. Babaguchi, “Classification-based Group Photo Retrieval with Bag of People Features,” Proc. of ACM International Conference on Multimedia Retrieval (ICMR2012), 8pages, June 2012 (査読有).
- [4] 中井優次郎, 新田直子, 馬場口登, “コミュニティに基づく画像コレクションの構造化”, 電子情報通信学会 2012 年総合大会, D-12-33, p. 127, March 2012.
- [5] K. Shimizu, N. Nitta, and N. Babaguchi, “Learning People Co-occurrence Relations by Using Relevance Feedback for Retrieving Group Photos,” Proc. of ACM International Conference on Multimedia Retrieval (ICMR2011), 8pages, April 2011 (査読有).

6. 研究組織

(1) 研究代表者

新田直子 (NITTA, Naoko)

大阪大学・大学院工学研究科・講師

研究者番号: 00379132