Vol.2010-DD-74 No.4
2010/1/29

自然文から視覚情報を復元する一手法と人物検索への応用

細見 格†

本稿では、アーカイブ画像の検索等での利用を想定し、自然文から画像特徴量を生成する手法について考察する。人が目で視た情報を言葉で記録した文書と、カメラで撮影した画像データとの関連付けにより、各々のメディアをより有効に活用できると考えている。今般、目撃証言などの日本語文から人物の服装に関する記述を抽出して画像特徴量に変換する手法を考案し、人物映像検索用のシステムにクエリ変換エンジンとして組み込むことにより、その有用性を確認した。

A Method to Reconstruct Visual Information from Natural Language Text and its Application to Person Search

Itaru Hosomi†

In this paper, we discuss how to generate image features from natural language phrases for retrieval of archived image data. We think that associating text data with image data both about visual image can improve value of these types of data. So we have developed a method to reconstruct visual information from natural language text such as description of witness accounts, and confirmed the method can contribute to human image retrieval through our prototyping.

1. はじめに

人間は、目で視た情報を言葉で表し、文書として記録したり他人に伝えたりすることができる。すなわち、視覚情報を言語情報に変換できる。一方、文書や音声で言語情報を受け取った人間の脳内では、その内容を視覚的記憶と結びつけて理解するために、抽象レベルで言語情報から視覚情報への逆変換が行なわれていると推測できる。この逆変換を計算機で再現できれば、文書と映像の意味的な統合管理や、言葉による映像検索などが可能になる。

筆者らは、人間が持つユニバーサルな情報伝達手段である自然言語を中心に据え、画像や音声などを含むメディア横断的な情報活用技術の研究開発に取り組んでいる。その一環として、現在、自然言語で表現された色や物体の構成に関する記述から画像検索用のデータを生成する手法を検討している。自然言語の語彙を用いた画像検索としては、Flickr や YouTube に代表される静止画／動画共有サイトのキーワード検索がよく知られている。最近では、Google などの Web 検索エンジンも、Web ページと共に画像を検索結果の一部として表示するようになっている。しかしながら、これらの検索は予め個々の画像データに結び付けられたテキストデータ（メタデータ）と入力キーワードとの文字列照合によって実現されており、キーワードが指す対象物の視覚的イメージを考慮した検索や、自然言語が持つ幅広い表現力を用いた感覚的な検索ができるわけではない。

言葉で表された情報から視覚的な情報（物理的なモノの姿）を網羅的に検索したい場面の例として、紛失物や危険物、所在不明の人物などの捜索が挙げられる。探したい人物としては、事件容疑者や迷子がある。人物捜索の場合、「紺のジャンパーを着た30 歳くらいの男性」のような言語情報から、店舗内の監視カメラなどで撮影された映像アーカイブを対象に検索するといったタスクになる。筆者らは、前述のような人物映像の検索を最初のターゲットとし、目撃証言等に含まれる服装の特徴記述を画像特徴量に変換することで検索可能にする手法を考案､プロトタイプシステムを試作した。

本稿は、以下の構成で進める。2 章では、言語情報から視覚情報を復元する手法と画像検索における位置付けについて考察し、既存の手法を概観する。3 章では本研究におけるアプローチと提案手法の概要を述べ、4 章で提案手法を用いた人物検索システムについて説明する。5 章では、試作したシステムを用いた言語情報から視覚情報への変換に関する予備的実験と評価結果について述べ、6 章で全体をまとめる。

† NEC 共通基盤ソフトウェア研究所
Common Platform Software Research Laboratories, NEC Corporation

ⓒ2010 Information Processing Society of Japan

2. 言語情報からの視覚情報の復元

2.1 人物映像検索を目的とした自然文からの視覚情報の復元

自然言語は膨大な語彙を有しており、その表現力は人間の五感によって知覚される殆どあらゆる現象について記述することができる。しかし、自然言語は個々の現象やそれらの要素を名詞や形容詞のラベルで記述するため、使うラベルが示す対象の粒度やラベル自体の曖昧さによって、その描写には厳密さを欠く場合が多い。それでも、どのようなタイプの物や情報であれ､それを探す際には自然言語が最もよく使われる。人に尋ねる場合はもちろん、検索ツールを使う場合も、キーワードやフレーズ、文を用いて問い合わせを行なう。写真や似顔絵、鼻歌など、検索対象と同じメディアタイプのクエリで問い合わせるリッチメディア検索もあるが、主流にはなっていない。

本研究では、自然言語による画像検索にフォーカスし、特に最初のターゲットとして人物映像の検索を選んだ。人物を映像データのアーカイブから探すという作業は、監視カメラ映像を対象に警察や警備員がよく行なう。その際、人物の視覚的特徴を表現するために、服装、身長、性別などが記述される。筆者らは、これらのうち服装の特徴記述に注目し、実際の視覚情報をどの程度復元しえるか、人物映像検索に実用たりえるかについて、検討と実験を行なっている。

2.2 画像検索のタイプと関連技術

記録映像データから特定の物体やシーンを探し出す作業は、一般に画像検索または映像検索と呼ばれる。本節では、より一般的な「画像検索」という呼び名を使う。画像検索の手段としては、キーワードまたはサンプル画像をクエリ（問合せ情報）とした静止画像や動画像の自動検索技術が実用化されてきている。

表 2-1 画像検索のタイプ

クエリ＼検索対象	テキスト（メタデータ）	画像
テキスト	メタデータによる画像検索	メディア変換型画像検索
画像	－	類似画像検索

画像検索には､表 2-1 のように幾つかのタイプがある。1 章で挙げたような Web 上の画像検索サービスは、その大半が予め個々の画像データに付与されたテキスト（メタデータ）をキーワードで検索する手法を用いており、画像検索ではあるが文字列照合によって実現されている。これは、表 2-1 においてクエリも照合対象もテキストの「メタデータによる画像検索」に相当する。[1] 一方、クエリも照合対象も画像データを用いるタイプは、「類似画像検索（画像内容に基づく検索）」に分類できる。[2) 3)]

これらに対し、テキストデータをクエリとして画像データと照合するタイプを、ここでは「メディア変換型検索」と呼ぶことにする。メディア変換型検索は、キーワードや自然言語文を記述したテキストデータを画像特徴量のデータに変換して、画像データから抽出した画像特徴量と照合することによって検索を実現する。[4) 5)] 実用化されている例としては、ビバコンピュータ株式会社が大阪府立大学と共同開発したデジタル画像検索エンジン「FeelPhrase®」（http://www.viva.co.jp/）があり、同技術を使った画像検索サイト「Feelimage®」（http://www.feelimage.jp/）が公開されている。同技術では､例えば｢青っぽいＴシャツ｣というクエリを入力すると､そこから色の名称｢青｣、色の修飾表現「っぽい」、対象物「Ｔシャツ」を抽出し、HSV 色空間（色相、彩度、明度の 3 軸からなる空間）で標準的な青色が占める領域に対して修飾表現に対応した補正（｢青っぽい｣では青の類似色を含んだ領域に拡張）を行なう。そして、補正後の領域内の色を主体とする画像を、対象物「Ｔシャツ」に分類された画像群から選び出す [6)]。この最後の処理が、画像特徴量同士の照合（ここでは形状特徴量を用いていないため、色彩特徴量のみの照合）になる。

なお、組合せとしては、画像データをクエリとしてテキストデータ（メタデータ）と照合するタイプもあり得るが、このような上記とは逆のメディア変換を用いた方式は例が見られず、画像検索という目的からも考え難い。

2.3 伝聞情報に基づく人物映像検索

人物映像を検索する現実的な手段として、次のような４つのタイプが考えられる。

① プロファイル情報と共にシステムに記録されている人物の画像を名前などで検索
② 対象となる人物のサンプル写真を元に、同じような顔や服装の人物画像を検索
③ 聞き込みなどで得た情報を元に、そこに現れている特徴を持つ人物画像を検索
④ 似顔絵やスケッチを元に、そこに現れている特徴を持つ人物画像を検索

①はメタデータによる画像検索の技術で、②は類似画像検索の技術でそれぞれ実現できる。③にはメディア変換型画像検索が適用できる。④は基本的に②と同様だが、写真に比べて似顔絵やスケッチは画像としての情報量が少なく一般的に精度も低いため、②に比べて難しい。筆者らは、③のタイプの検索に注目し、目撃者から聞いた伝聞情報に基づいて人物を探すための検索における課題を検討した。

③のタイプの検索に使える技術としては前述した杉浦 [4)]や三木 [5)]の方式があるが、これらは１クエリにつき基本的に１色のみの指定を想定している。三木 [5)]の技術を用いた画像検索サイト Feelimage®で複数色を指定した場合は、単純な OR 検索（いずれかの色を含む画像を検索）となる。すなわち、複数色を指定することによって検索対象の画像を適切に絞り込むことができない。現実世界では、ある対象物（風景なども含む）の表面が２色以上または２系統以上の色で構成される場合も多く、特に形状の特徴で絞り込めない場合には、複数色の組合せを絞り込み条件として利用できることが重要と考えられる。そこで、本研究では、

ⓒ2010 Information Processing Society of Japan

*自然言語で記述されたクエリから画像を検索する技術において、複数の色を指定したクエリからでも検索対象画像を適切に絞り込める画像特徴量を生成すること*を主たる課題としている。

3. アプローチ

3.1 要件の分析

人物の映像（静止画像または動画像）を検索する場合、検索対象となる映像の品質や数量（バリエーション）がまず問題となる。証明写真のように良好な照明条件の下で、カメラに対して正面を向いている人の映像や、ハイビジョンカメラで撮影された高解像度の映像であれば、最新の画像処理技術を用いればかなり豊富で正確な情報を得ることができる。そうした人物カタログの検索のような場合に比べ、これまで述べてきたような監視カメラ映像を検索対象とする場合、解像度、照明、映っている人の位置や向きなどの条件が相当厳しくなる。多くの監視カメラ映像は 640x480 ピクセル程度の画素数で、しかも映っている人物は画面内のごく一部である場合が多い。人物の向きもカメラの方を向いていることの方が稀であり、多くは上方から見下ろした視点になっている。撮影場所がやや暗かったり照明の種類の違いによって、実際とはやや異なる色に映っていることもある。こうした条件下では、撮影された対象の形状や模様を正確に判別することは困難と考えられる。一方で、色の識別については、設置されたカメラ毎に予めホワイトバランスなどの調整を施しておけば、形状よりも判別し易いと考えられる。

次に、人の服装が自然言語でどのように表現されるかについて考察する。人物映像の検索を目的とした場合、警察が捜索協力の依頼用に公開している情報[6)7)]などから、自然言語で記される服装に関する検索条件は服の色と種類が主体になると考えられる。以下では、このような公開情報を参考に考察した結果を述べる。

まず色について、人が目で見たものの色を言葉で表現する場合、全体が単色または同系色であれば、その色に最も相応しいと思える色の名称を使う。ちょうど合致する名称が無ければ、最も近いと思われる色を表す名称に修飾語句を付与して表現することがある。その場合の修飾語句には、「明るい」「淡い」といった前置修飾語や「っぽい」「系」といった後置修飾語を使い、さらには「黒っぽい」「黄ばんだ」のように別の色の名称を使って修飾する場合もある。また、人が見て複数種類の色が混在していると認識した場合には、それらの色についての記述を並立助詞などの接続語句で結びつけた言葉で表現することが多い。その場合、構成色のうちいずれが主色であるかを明示する表現と明示しない表現があり、前者には並立助詞「に」（例：「白地に赤」）が、後者には「と」「や」など（例：「赤とピンク」）がよく使われる。主色は、他の付加的な色に比べて対象領域を占める面積比が十分大きいと考えられるため、適切な比率値を与えれば精度良く目的の画像を検索できる。

複数色の構成が個々の色で表現し難い場合や、ある色の組合せが一言で表現される場合がある。例えば「迷彩色」は、構成する色が必ずしも明確ではないが、緑や茶系の色、白などある程度限定された数色で構成されている。

服の種類については、ワンピースなど全身を覆うものの他、上半身用の上衣と下半身用の下衣に大別できるが、それぞれ複雑な柄や装飾の少ない一着の服であれば、前節のような色の自然言語表現から画像特徴量を決定することで、検索が可能になる。しかし、特に上衣の場合、シャツとジャケットなど複数の服の重ね着はごく一般的であり、且つどちらの服を上に着ているかについて、言葉で直接語られることは少ない。上半身で重ね着している場合、各服の色が画面に映っている面全体に占める割合が、色に基づく検索において問題になる。

例えば、クエリに「白いワイシャツと黒のジャケット、紺のジーパン」と書かれていた場合、各服の色の割合を決定するためには、どの服同士が重なり合い、どちらが上（外側）にくるのか、下（内側）になった服は外からどの程度見えているのかが分からなければならない。前記のクエリにはこれらを示す直接の情報は無く、服とその色の名称のみが記述されている。このような文を元に該当する人物を探すケースは、前述した警察の公開情報などから一般的であることが分かっている。

重ね着の場合には、さらに人の身体がどういった角度で画面に映っているかも、各色が占める割合に関係してくる。多くの場合、上着は身体の後面や側面を覆うが、正面では幾らか内側の服が見えている。内側の服が見えていなければ、そもそもクエリ（目撃証言など）に現れないはずである。従って、検索対象映像から人の向きが判定できるならば、人の正面映像のシーンでは、内側の服の色をある程度検索条件に含むといった工夫が必要になる。

以上の考察から、自然文をクエリとした人物映像の検索における要件を次のように整理した。

(1) 色の名称についての十分な知識を備えていること
(2) 色の修飾表現についての十分な知識を備えていること
(3) 色同士の関係を表す接続表現についての十分な知識を備えていること
(4) 検索したい対象が複数の色を含む場合に各色が占める面積比率を判定できること
(5) 服の名称についての十分な知識を備えていること
(6) 名称で特定された各種の服について、各々が覆う人の身体部位との対応関係を判定できること
(7) 複数の服を重ねて着ている場合に、その上下（外側と内側）関係や内側の服がどの面でどの程度外部から見えているかを判定できること

ⓒ2010 Information Processing Society of Japan

3.2 課題の具体化

前節の考察から、本研究の技術的課題を表 3-1 のような部分課題に分解した。

表 3-1 本研究の部分課題

No.	部分課題	例（矢印の右側は対応付けの例）
1	色の名称と色空間上の値との対応付け	青→{H:240, S:1.0, V:125}
2	色の修飾表現と色補正操作との対応付け	明るい→{H:0, S: -20%, V: +20%}
3	色の分配表現と配分比率との対応付け	白に青→白 80%, 青 20%
4	模様の名称と色の配分比率との対応付け	青と白のチェック→青,白各 50%
5	非明示的な色表現に対する色の配分の決定	迷彩色→緑 50%, 白 30%, 茶 20%
6	対象物の名称と身体部位との対応付け	ジャケット→上半身
7	指定した色と対象物との対応付け	緑のシャツと白のズボン
8	複数対象物の重なりによる色分配比率決定	黒のシャツにグレーのジャケット

表 3-1 の項目は、自然言語による対象物の配色に関する直接表現（No.1～No.3）、間接表現（No.4～No.5）、物体構成による表現（No.6～No.8、ただし No.6 は色自体とは直接関係しない）の 3 種類に分類できる。直接表現のうち、No.1 と No.2 については従来手法でも対応しているものがある。[4) 5)] また、No.5 のような色について非明示的な表現を実際の色に対応付けることも、比喩表現（例えば「秋色」）と同様に従来のヒューリスティックスや機械学習によるアプローチである程度対応可能と考えられる。No.6 と No.7 は、色や対象物に関する辞書と簡単な言語解析によって実現できる。しかし、No.3, No.4 のように複数色の配分や模様、No.8 のように物体の構成に依存した配色については、有効な従来手法が見つかっていない。

3.3 色の表現に関する語彙と服装構成知識を用いた画像特徴量の生成

上記課題への対応方針として、表 3-1 の右側に記載した例のように、まずは色や服の名称、色を修飾する表現、色同士の関係を表す表現に関する辞書を構築し、入力された自然文を語の列として解析する。形態素解析や係り受け解析を使ったより厳密な方法も採り得るが、前述のような名称や修飾表現とそれらの並びからかなり判断できると推測した。また、各色の名称に対応する色空間上の値は予め辞書に登録しておく。

服と身体部位との対応付けや服同士の重ね着への対応には、人の身体構成（上半身/下半身、正面/非正面）と服との対応関係、および服のタイプ（上着/下着）をクラスとし、その間の関係と共にオントロジを定義して「服装構成知識」を構築することで、服装から各服の色の見え方を決定することとした。

4. 人物検索システム

4.1 クエリ変換エンジン

前節の方針に基づき、自然言語文を用いて人物画像（動画像を含む）を検索するために必要な「クエリ変換エンジン」を設計した。ここで言うクエリ変換エンジンとは、図 4-1 のように、自然言語文を入力とし、識別対象の構成単位（図では「上半身」と「下半身」）ごとの色に関する画像特徴量を出力するソフトウェアモジュールである。

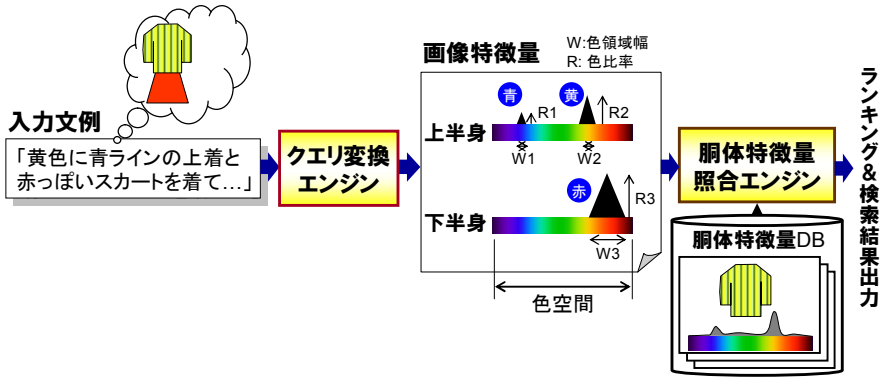


図 4-1 クエリ変換エンジンの入出力イメージと人物検索における位置付け

クエリ変換エンジンが出力する画像特徴量は、画像上の各構成単位を表す領域内の主要な色について、その画像上および特定の色空間上での分布を表したものである。図 4-1 に示したイメージの例では、1 次元の色相空間上で「青」、「黄」、「赤」という色名が対応するそれぞれ W1～W3 の幅を持った部分領域と、前記した 1 つの構成単位を表す画像上の領域で個々の色が占める面積比率を表す値 R1～R3 の組を、「上半身」と「下半身」それぞれについて記述したものを出力している。図では色空間を 1 次元の色相のみで表しているが、実際の画像特徴量には 3 次元の HSV（Hue：色相、Saturation：彩度、Value：明度）色空間を用いている。さらに、H、S、V の各値に対して色名毎に一定の幅 W（図では W1～W3）を与えることにより、言葉で指定された色の曖昧さに対応している。

出力した画像特徴量は、別途開発中の「胴体特徴量照合エンジン」によって、予め蓄積された人物映像から抽出した大量の胴体特徴量（画像認識技術で自動抽出した人物の胴体領域の画像特徴量）と照合し、類似性の高い胴体特徴量を持つ人物の映像を検索結果として返す。

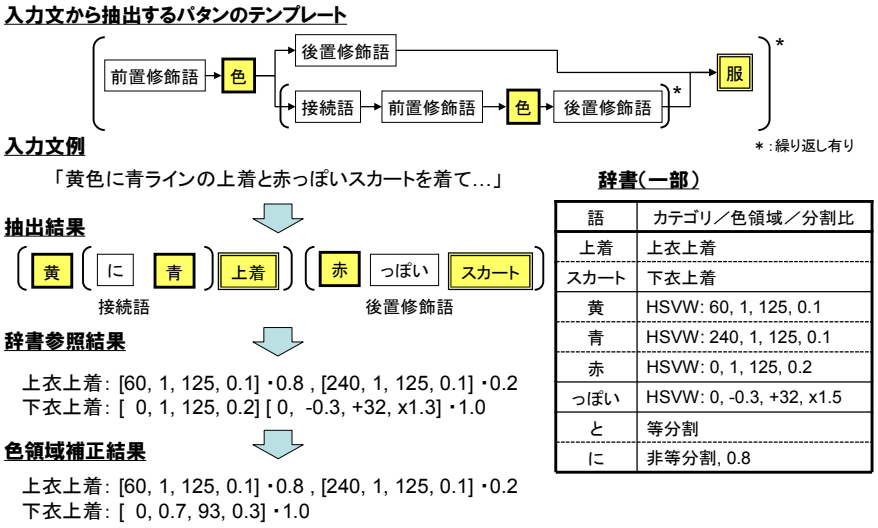
ⓒ2010 Information Processing Society of Japan

クエリ変換エンジンでは、前節の方針により、画像上の個々の識別対象領域（ここでは１人の人物の足元から首までの胴体領域）を占める色の値（色空間上の部分領域）とその面積比率を算出するために、ヒューリスティックな自然言語解析とオントロジに基づく服装構成知識を利用している。以下、これらの具体的な手法を説明する。

4.2 色空間上の領域と色面積比率の判定

ある領域が複数の色で占められていた場合、その領域の色を表現する自然言語の語彙として、「色の名称」と「色の修飾語句」、これらを使った各色についての「色記述間の接続語句」をそれぞれ辞書化して入力文を解析することにより、各色の色空間上での値と占有面積の比率を決定する。色の名称と色空間上の値との対応関係は、Wikipedia に記載された HSV 値を参考に、適宜補正して辞書化した。修飾語句の辞書化についても、個々の修飾語による色空間上の値の変化量をヒューリスティックに与えた。接続語句については、例えば接続語「に」の場合、直前の色を主色とし、直後の色を部分色として、面積比率を主色の方に大きく割り当てた。主色が明確でない「と」などで接続された色同士は、面積比率も同程度と見なす。

以上のような自然言語表現についてのヒューリスティックな辞書を利用することで、図 4-2 の例のように文章から色に関する記述を抽出し、色空間上の値と色同士の面積比率を決定する。



語	カテゴリ／色領域／分割比
上着	上衣上着
スカート	下衣上着
黄	HSVW: 60, 1, 125, 0.1
青	HSVW: 240, 1, 125, 0.1
赤	HSVW: 0, 1, 125, 0.2
っぽい	HSVW: 0, -0.3, +32, x1.5
と	等分割
に	非等分割, 0.8

図 4-2 色に関する記述の抽出と辞書の参照による画像特徴量の決定方法

図 4-2 の例について説明する。入力文「黄色に青ラインの･･･」が与えられると、図の上部のようなテンプレートを用いて必要な要素を対象物の構成単位毎（ここでは服単位）に抽出する。抽出した各要素について辞書を参照し、構成単位についてはカテゴリ（「スカート」のカテゴリは「下衣上着」）、色の名称については色空間上の領域（中心の HSV 座標値と幅 W）、色の修飾語については前記領域に対する値操作係数、色記述間の接続語についてはその前方の色と後方の色との面積比率（分割比率）を取得し、それぞれ抽出結果に反映する。さらに、各色の色空間上の領域を接続語の係数で補正した結果を画像特徴量とする。図中の “[60, 1, 125, 0.1]・0.8” といった記述は、[ ] 内が HSV 色空間の座標 H（色相）: 0～360°, S（彩度）: 0～1.0, V（明度）: 0～255 の各軸の値に対して幅 W（全幅に対する割合）: 0～1.0 を半径とした球状の色領域を表し、その色領域が構成単位の全領域に占める面積比率を “・0.8” で表している。

前述した警察署が公開していた 2009 年 7 月 30 日時点の計 101 件の服装情報では、うち 8 件が複数色の組合せであった。[7) 8)] 「紺色に白色ライン」のような主色＋部分色を指定する場合に並立助詞「に」が使われ、その他は並立助詞「と」の使用が多い。また、全体としては、目撃者の記憶に基づく証言が大半であるためか、「っぽい」（例：「黒っぽい」）や「または」（例：「水玉または緑色」）といった曖昧な表現が目立つ。ただし、人の視覚による色の認識と言葉による表現は、その場の照明条件や周囲の色、精神状態や語彙知識などにも影響されるため、厳密な対応付けは本来困難であることに留意する必要がある。

4.3 服装構成知識を用いた色面積比率の決定

服の名称と人の身体部位（全身/上半身/下半身）および重ね着時の上下位置との対応関係を「服装構成知識」として持たせ、重ね着時に上になる服の色と下になる服の色との面積比を固定値で与えることとした。ただし、人体正面を写した画像とは設定した面積比で照合するが、正面以外（側面や背面）では上になる服の色のみと照合する。このような照合ルールは図 4-1 の胴体特徴量照合エンジンで適用するため、クエリ変換エンジンは正面用と非正面用でそれぞれ別の画像特徴量を出力すれば良い。

なお、検索文に記載された服の名称によっては、それが全身を覆うものか上半身または下半身のみを覆うものかが判断し難い場合がある。例えば、「ジャージ」や単に「服」といった場合である。こうした場合、同文に身体部位との対応が明確な別の服の名称があれば、そちらを優先する。これは、身体部位不明または全身用の服を「衣服」カテゴリとし、「上衣」や「下衣」などに属する服は「衣服」に属する服に優先する（後述の「上書き優先」）というヒューリスティック・ルールで対応している。

服装構成知識の概略を図 4-3 に示す。

ⓒ2010 Information Processing Society of Japan

Vol.2010-DD-74 No.4
2010/1/29

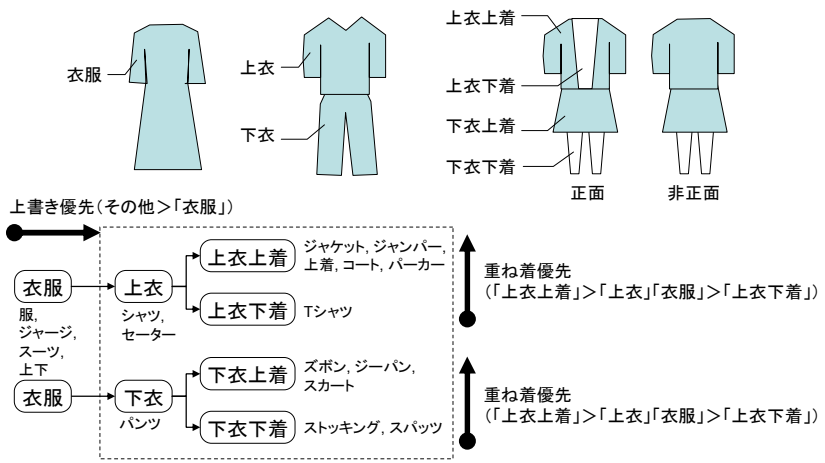


図 4-3 服装構成知識の概略

服装構成知識は、小規模なオントロジで定義している。ここで言うオントロジとは、ある事物を分類するクラス（カテゴリ）や事物間の関係、属性などを表す語彙、およびそれらを定義したものを指す。服装構成知識については、「衣服」「上衣」「上衣上着」などをクラスの語彙、「上書き優先（上書）」「重ね着優先（重着）」を関係の語彙としたオントロジを用いる（図 4-4）。

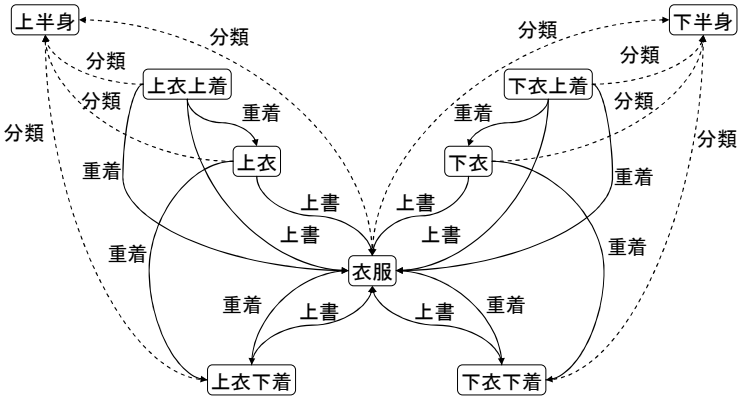


図 4-4 オントロジに基づく服装構成知識

なお、一般的オントロジには必ずある「サブクラス」という関係を用いることで、「上衣」は「衣服」のサブクラス、「上衣上着」は「上衣」のサブクラスであると定義できるが、現状のクエリ変換エンジンではそのような上位下位関係を利用していない。

4.4 人物検索システム

以上の実現方法に基づき、自然文から色彩特徴量を主体とした画像特徴量データを生成するクエリ変換エンジンを試作し、監視カメラ映像アーカイブを対象とした人物検索システムに組み込んだ。人物検索システムは、特定の人物が写っている映像を、服装画像または顔画像の照合によって検索できる。服装画像の照合による検索では、クエリ変換エンジンを用いた自然言語文からの検索に加え、服装のサンプル画像や顔画像からの検索にも対応している（図 4-5）.

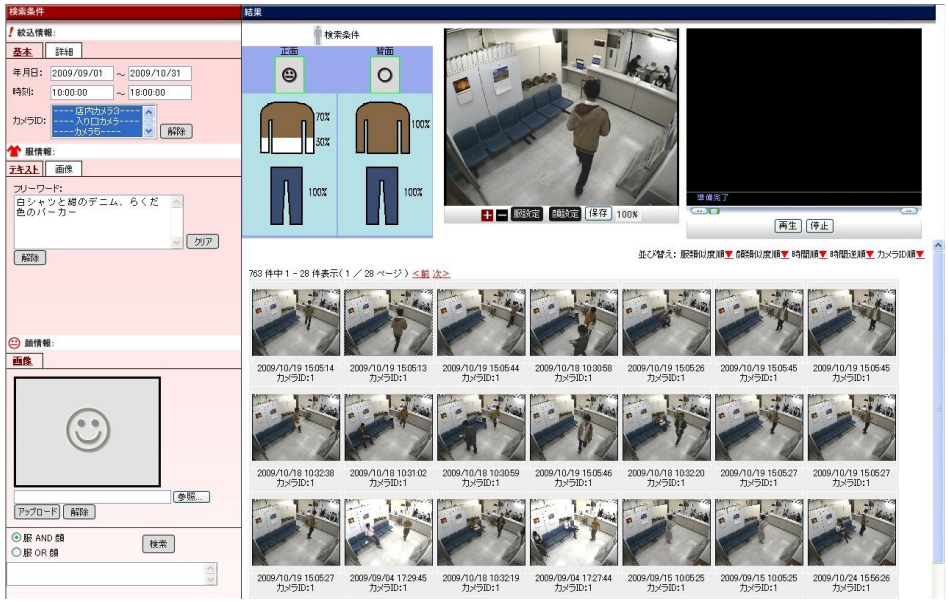


図 4-5 人物検索システムの画面と検索例

試作した人物検索システムでは、画面左側中段にフリーワード入力欄があり、ここに自然文でクエリを記入できる。クエリを記入して「検索」ボタンを押すと、その直ぐ右上にある「検索条件」欄にクエリの解釈結果として人の上半身／下半身、正面／非正面それぞれの色と面積比率が表示される。同時に、検索結果のプレビュー画像一

ⓒ2010 Information Processing Society of Japan

覧が右下に表示される。プレビュー画像のうちいずれか 1 つをクリックすることにより、画面右上で拡大表示や動画の再生ができ、検索結果の詳細を確認できる。

図 4-5 の例では、クエリとして「白シャツと紺のデニム、らくだ色のパーカー」と入力し、画像特徴量として上半身正面がらくだ色（やや濃い黄土色）70%と白 30%、下半身正面が紺色 100%、上半身非正面がらくだ色 100%、下半身非正面が紺色 100%に変換されている。また、実際に検索結果一覧の多くに同じ人物が映っている。入力されたクエリについて、シャツとパーカーはそれぞれ上衣であり、デニムは特に上衣を表す語と結びつかない場合ジーンズと同様に下衣と判定している。また、パーカーは上着、シャツは下着に分類されるため、パーカーの色が上半身の大半（正面では 70%、それ以外では 100%）、シャツの色が上半身正面の 30%のみを占めている。これらの判断は、前述の服装構成知識に基づいたものである。

なお、クエリ変換エンジンから出力した画像検索用データは、さらに上/下半身・正面/非正面毎の HSV 色空間上のヒストグラムに変換して蓄積画像の特徴データと照合しているが、本稿では、その具体的なアルゴリズムについての説明は割愛する。

5. 実験と評価

5.1 実験

実験に用いた辞書には、63 の服名、40 の色名、11 種の色修飾語（「明るい」「淡い」など）、8 種の色記述間接続語（「と」「や」「、」など）を登録した。実験用のクエリには、まずリファレンスデータとして 4.2 節で述べた警察が公開している 101 件の不審者情報を用いた。また、20 種類の服装を着用した人物を 3 台のカメラで異なる場所と角度から撮影し、複数の人にそれぞれの服装を言葉で表現して貰うことで、クエリ変換および人物映像検索の評価用クエリ 25 件（服装の重複あり）と、対応する人物の映像データ約 1.8 時間分を用意した。ただし、本稿ではクエリ変換の実験と評価のみについて述べる。

本実験において、正解の判定は、クエリ変換エンジンの出力結果（図 4-5 に示した人物検索システム画面上部の「検索条件」欄に表示される身体各部位の色と面積の割合）が人間によるクエリの解釈と合致しているか否かで行なった。今回試作したクエリ変換の精度は表 5-1 のような結果となった。

表 5-1　クエリ変換の精度評価結果

クエリの種類	クエリ総数(A)	正解数(B)	正解率(B/A)
警察署公開文	101	91	90.0%
独自収集文	25	14	56.0%

5.2 評価

警察署公開文は、もともと本クエリ変換エンジンの設計において参考にしたデータであるため、高い精度が得られている。また、捜索の協力を仰ぐ公開情報であるため、誰が読んでも一意に解釈できるように、なるべく明確かつ一定した様式で目撃情報が記載されていることも、高精度の理由になっていると思われる。一方、筆者らが独自に収集した文は、各自が目の前の人物の服装について自由に表現した文言ほぼそのままであるため、総数が少ないにも関わらず表現のバリエーションが多い。それらのバリエーションは、警察署公開文を参考に設計した現在のクエリ変換エンジンではカバーできないものが多く、それが精度に現れたと考えられる。

正しく変換できなかった 11 件のうち、2 件は辞書の不足、2 件は自然言語解析機能の不足、1 件は上着同士の重ね着（ジャケットとパーカー）で服装構成知識から色面積比率を判定できなかったケースである。残り 6 件を下記 4 種の課題に整理した。

(1) 服の重ね着判定精度の向上（例：「青のスーツに白いシャツ」）
(2) 同色服の色指定省略への対応（例：「紺のベストと（紺の）スカート」）
(3) 他の色による修飾と複数色の組合せとの区別（例：「赤紫」「白黒」）
(4) 付帯物や部分的衣服への対応（例：「赤のネクタイ」「茶色の短パンに白の靴下」）

(1)に関して、上下一体の服である「スーツ」に対し、上半身ではその下に着用する「シャツ」との関係が正しく判定できず、実験では上半身が白 100%となった。これは服装構成知識における「上書き優先」の定義に問題があったことを示している。(2)は、上下同色の服の場合、後に述べた服の色を省略されるケースへの対応である。先に述べた服の色と同色なのか、単に印象に残らず省略されたのかは一概に判定できないが、可能性として前出の色を継承しておくことは有用と思われる。(3)については、「赤紫」は「赤っぽい紫」であり色相を赤の方へシフトする操作が適切だが、「白黒」は「白っぽい黒」ではなく「白と黒」の組合せと解釈する必要がある。(4)は、アクセサリや鞄などを服装の一部として取り込むことと、半袖の服や短い下衣など肌や下着の色が部分的に見えるケースへの対応を指している。これらの課題を解決し、辞書を拡充することで、今回の独自収集文のような発言者の表現そのままのクエリからでも、警察署公開文と同程度の精度が得られると見積もっている。

なお、自然言語解析を簡易な手法で行なっている点の限界も見えている。独自収集文の 1 つに「黄色の T シャツの上に赤っぽい柄シャツ」という表現があり、この中で「～の上に」が正しく解釈できず「に」だけを取って「黄色の T シャツに赤っぽい柄シャツ」と同じ解釈結果になった。すなわち、上衣はいずれも「シャツ」であるため服の種類では重ね着の上下が判断できず、接続語「に」から上半身の主色は黄色となった。実際は、「柄シャツ」の方が上着であるため主色は赤っぽい色となるべきであるが、その正しい判定には「の上に」を辞書に登録するか、より詳細な言語解析が必要となる。

ⓒ2010 Information Processing Society of Japan

Vol.2010-DD-74 No.4
2010/1/29

6. おわりに

本稿では、自然言語による言語情報から視覚情報を復元することの可能性と手法について考察し、自然文から視覚情報に相当する記述を抽出して画像特徴量に変換することで、人物映像の簡便な検索を実現する手法と試作したシステムについて述べた。また、試作システムを用いた予備的実験により、本提案によるシンプルな手法でも警察署が公開しているような比較的フォーマルな文章からは高い精度で視覚情報を復元できること、逆に自由な発言への対応にはまだ課題が多いことを確認した。現在、5.2節で挙げた課題(1)～(4)に対応した改良版の開発に取り組んでいる。

現時点は、まだ研究の初期段階であり、考察の網羅性やクエリ変換エンジンの機能、辞書のチューニングなどいずれも不十分である。実験データの拡充も大きな課題であり、技術改良と併せて進めていく。

参考文献

1) 井原, 林, 尾内, "もぶろげっと：画像情報を含む Blog 記事検索システム", インタラクティブシステムとソフトウェアに関するワークショップ（WISS2005）論文集, pp.69-73, 2005.
2) 小早川 他, "ウェーブレット変換を用いた対話的類似画像検索と民俗資料データベースへの適用", 情処学会論文誌, Vol.40, No.3, pp.899-911, 1999.
3) 北本, 高木, "類似画像検索への応用を目的とした階層化属性付きグラフマッチングの高速化", 画像の認識・理解シンポジウム (MIRU'96), Vol.II, pp.331-336, 1996.
4) 杉浦, 原田, 中谷, 伊東, "自然言語による画像データベース検索", 信学技報, PRU94-13, pp.55-62, 1994.
5) 佐藤, 山崎, “服装・顔を用いた人物検索システム”, Matsushita Technical Journal, Vol.52, No.3, pp.67-71, 2006.
6) 三木, 馬野, "画像蓄積・検索システム及び画像蓄積・検索システム用プログラム", 特開 2009-3581（特願 2007-161858）, 2007.
7) 警察署別子ども対象犯罪声かけ事案一覧, 平成 21 年 7 月 30 日時点, http://www.pref.nagano.jp/police/seian/shonen/kodomo_anzen/jian_ichiran.htm
8) 松山西警察署管内にこえる不審者等の情報, 平成 21 年 7 月 30 日時点, http://www.police.pref.ehime.jp/nishi/fushinsya.htm

ⓒ2010 Information Processing Society of Japan