

あいまいな属性を用いた人物画像検索

董 テイ テイ[†] 吉田 登[†] 劉 健全[†]

[†] NEC バイオメトリクス研究所 〒 211-8666 神奈川県川崎市中原区下沼部 1753

E-mail: †{t-dong,n-yoshida14,jqliu}@nec.com

あらまし 属性に基づいた人物画像検索は、画像から検出した人物の属性に対し、指定した属性条件に満たしている人物画像を検索する。目撃情報による不審者捜査や迷子捜索、ファッショントレンド分析など幅広い応用がある。ユーザは記憶から対象人物の特徴を思い出す必要があるため、指定した属性条件には固有のあいまい性または不確実性がある。また一方で、既存の画像を用いた属性推定技術では、一般的に属性とともに確信度スコアが推定される。従来研究では、属性の確信度を利用しないか、閾値で確信度の高い属性を抽出して使用するため、確信度の低い属性は検索に利用されない。また、ユーザが指定する属性の不確実性も考慮されない。そこで本研究では、属性条件のあいまい性を確信度で表現し、ユーザが属性条件の確信度を指定できるシステムを提案する。これにより、属性条件に内在するあいまい性と画像からの属性推定に伴う不確実性の両方を考慮して、あいまいな属性を用いた人物画像検索を実現する。また、確率的データベースに関する知識を活用し、不確実性をそれぞれ含む属性条件と人物属性との類似度を計量可能な、新たなマッチングスコアを定義する。さらに、確信度臨界点という新たな定義を提案し、各属性条件の確信度が検索結果のランキング順位に与えるインパクトを示す。また、ウェブ上で動作するデモシステムを開発し、防犯カメラの実映像データを用いて、本研究の有用性を示す。


キーワード 人物検索, 画像, 属性, あいまい, 不確実, 確信度

1 はじめに

属性に基づいた人物画像検索は、年齢や性別、服の色などの属性条件を指定し、画像または映像から対象人物を検索する [2], [5], [8]。フォレンジック分析や映像監視、ファッショントレンド分析などの分野で幅広い応用がある。例えば、目撃情報に基づいて公共の場で現れた不審者や迷子を検索する等である。

属性に基づいた人物画像検索は、実世界で対象人物の画像が入手できない場合は特に有用である。一つの解決策として、自由形式テキストまたは自然言語表現を用いて検索する方法がある [7], [11]。しかしながらこの方法は、自然言語固有のあいまいさにより同一の状態を指す表現が一つに絞られないケース（赤い服、赤色の服、赤の色をした服等）が多く、検索条件を解析することが複雑であるため適用が難しい。それに対して、画像認識技術によって推定される属性（年代、服装、服色など）はあいまいさが少なく、さらに認識技術の進展により収集しやすくなってきていることから、適用が容易である [10], [15]。

問合せ画像で検索する人物画像検索 [3], [9], [12], [16] に比べて、属性に基づいた人物画像検索は柔軟性がより高いが、ユーザが適切に検索に必要な属性条件を指定する必要がある。一方、ユーザはほとんどの場合記憶から対象人物の特徴を思い出すため、指定した属性条件があいまいになることがある。また、光などの環境条件や視点等の変化、隠れの有無などもあいまい性が生じる要因として挙げられる。このようにユーザが指定した属性条件に固有のあいまい性または不確実性が含まれると、ユーザが指定した検索条件と、ユーザが本当に探したい属性情



Person Attribute	o_1	o_2	o_3	o_4
Personal Less30	0.3912	0.1457	0.2502	0.9309
upperBody Pink	0.9985	0.9962	0.5020	0.5117
lowerBody Black	1.0000	1.0000	0.3752	1.0000

図 1 オブジェクト検出技術 [13] で検出された人物と、属性推定技術 [10] で推定された属性と確信度スコアの例。

報とにギャップが生じ、望む結果を得ることが難しくなる。結果として、ユーザは対象人物を見つけるために何回も条件を変更しながら検索を試さなければならなくなる。

また一方で、既存の画像を用いた属性推定技術では得られる属性情報も不確実性を含んでおり、不確実性は推定した属性に対する確信度スコアとして出力される。すなわち確信度スコアとは、推定した属性がどれぐらい信頼できるかといった度合いを示す値である。図 1 はオブジェクト検出技術 [13] で検出された人物と、属性推定技術 [10] で推定された属性と確信度スコアの例を示している。しかしながら、多くの属性を用いた人物画像検索に関する従来技術では確信度スコアを利用しないか、閾値処理で確信度スコアの低い属性を取り除いているにすぎず、確信度スコアが有効に活用されていない。

以上の背景から本研究では、ユーザが検索時に指定した属性条件が含むあいまい性と、画像から推定された属性の不確実性の両方を有効に活用した検索技術を提案し、属性を用いた人物

画像検索の改善を目指す。具体的には、ユーザが各属性条件に対する不確実性を確信度として入力できるようにすることであいまいさを反映し、さらにデータベース内の画像に対して推定された属性の確信度も加味して、確率的データベース [14] の知識を用いてあいまいな属性を用いた人物画像検索を実現する。

本技術の実現に向け、属性のあいまい性または不確実性を考慮することができる新たなマッチングスコア、及び本手法を取り込んだ検索システムを提案する。また、ユーザが指定する確信度が検索結果に与えるインパクトを可視化するために、**確信度臨界点**という新たな定義を提案する。確信度臨界点とは、検索結果のランキング順位が入れ替わる確信度の値を表す。

本技術を実装したデモシステムを開発し、防犯カメラの実映像データでテストを行った。本デモシステムはウェブ上に構築されており、ユーザはわかりやすいユーザーインターフェースを用いて映像をアップロードし、属性推定と人物画像検索を行うことができる。

2 用語定義

本論文では、画像中の各人物に対して、属性及び確信度スコアが推定されている状況を想定する。人物検出には“*You Only Look Once (YOLO)*” [13] などを用いることができ、属性推定には例えば、深層学習に基づいた複数属性推定技術である *DeepMAR* [10] などを用いることができる。

本研究では、以下のように問合せとのマッチングスコアを定義する。問合せ q は属性の特徴量と確信度のペアの集合 $\{(f_j^q, p_j^q) | j \in [1, m]\}$ で表現する。 m はユーザが指定した属性条件の数である。人物オブジェクト o と問合せ q との**マッチングスコア**は、 $S(o, q)$ で表し、以下のように定義する：

$$S(o, q) = \sum_{j=1}^m p_j^o \times p_j^q \times \text{Sim}(f_j^o, f_j^q) \times w_j. \quad (1)$$

ただし、 f_j^o と p_j^o は o の属性の特徴量と確信度、 $\text{Sim}(f_j^o, f_j^q)$ は o と q それぞれの属性に関する特徴量間の類似度、 w_j は j 番目の属性の重みである。二つの属性に関する特徴量間の類似度を求める関数は、属性の表現に使われる特徴量に依存するが、コサイン類似度などよく知られている類似度関数で計算できる。

属性の重みは各属性の類似度を最終的な類似度にどの程度反映させるかを示す値であり、以下の通り属性の識別能力に基づいて算出する。まず、属性に関する類似度で人物画像をクラスタリングする。具体的には、例えばある閾値で属性に関する類似度が閾値を超えている人物画像が同じクラスに属するようにする。その後、以下の式により各属性の重みを計算する：

$$w_j = \frac{1}{N_C} \sum_{r=1}^{N_C} \frac{1}{|C_r|}. \quad (2)$$

ただし、 N_C は各属性におけるクラスタの数であり、 $|C_r|$ は r 番目クラスタ C_r の人物の数である。

問合せ q とデータセット D が与えられた際、本研究は D から q に対してマッチングスコアが最も高い上位 k 件の人物画像を検索する。

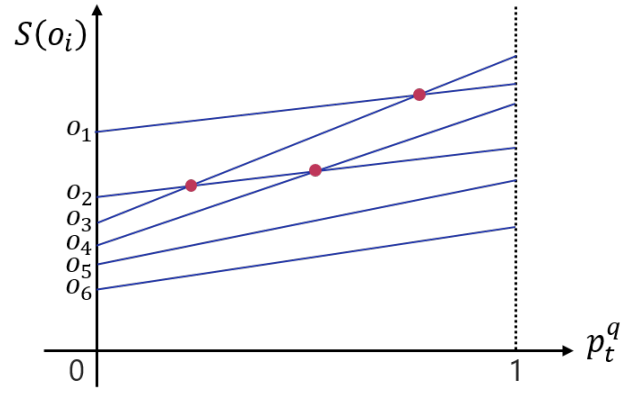


図2 確信度臨界点の例。確信度臨界点は二つのマッチングスコア線が交差している p_t^q の値である。この図では、三つの確信度臨界点があることになる。

3 提案手法

本研究では、まず確信度は考慮せずに指定された属性条件のみを用いて候補となる人物を抽出し、その後各人物候補に対して確信度を考慮したマッチングスコアを計算してランキングし、上位 k 件の人物結果を出力する。

3.1 確信度臨界点

式 (1) により、問合せ属性の確信度が変更された場合、すべての人物候補とのマッチングスコアが変わり、検索結果のランキング順位も変わる。よってユーザの入力する属性情報があいまいであった場合、ユーザは問合せ属性の確信度を変化させながら検索結果を確認する試行錯誤を繰り返して、適切な確信度を決める必要がある。しかし、試行錯誤の繰り返しには時間がかかる。

この問題を解決するために、**確信度臨界点**を導入する。確信度臨界点は、検索の問合せに入力するある属性の確信度を変化させた場合に、検索結果のランキング順位が入れ替わる確信度の値を表す。各問合せ属性に対して確信度臨界点を可視化することで、ユーザは実際に確信度の変更を試さなくても、その変更に伴って検索結果のランキング順位が変わるか否かを知ることができ、適切な確信度を容易に決定することができる。これにより、試行錯誤の回数を減らすことができ、探したい人物画像を絞り込むための検索時間を削減することができる。

図2は確信度臨界点の例を示している。人物画像 o_i の問合せ q に対するマッチングスコア $S(o_i)$ は、 t 番目問合せ属性の確信度 p_t^q に対して線形に変化する。二つ人物画像に対するマッチングスコアの線が交差している p_t^q が、確信度臨界点である。この図では、三つの確信度臨界点があることになる。

各問合せ属性の確信度臨界点を効率的に計算するために、Bentley-Ottmann アルゴリズム [4] を拡張し、マッチングスコア線が交差する人物画像のペアに対してのみ交点を計算した。具体的には、各属性において $p_t^q = 0$ と $p_t^q = 1$ でそれぞれすべての人物に対してマッチングスコアの高い順にソートする。例えば、

図 2 においてソート結果は、それぞれ $L = \{o_1, o_2, o_3, o_4, o_5, o_6\}$ と $R = \{o_3, o_1, o_4, o_2, o_5, o_6\}$ である。

交点を持つ人物画像のペアは、 L と R でランキングの序列が変化しているオブジェクトのペアのみである。

例えば、 o_1 と o_3 のペアに着目する。 L における序列は $\{o_1, o_3\}$ の順であるが、 R では $\{o_3, o_1\}$ の順に変化している。そのため、 o_1 と o_3 のマッチングスコア線が交差していることがわかる。次に o_1 と o_2 のペアに着目する。 L における序列は $\{o_1, o_2\}$ の順であり、 R でも変化していない。従って、 o_1 と o_2 は交差しないことが分かる。

このように、マッチングスコア線が交差しているオブジェクトのペアのみに対して計算することで、効率的に確信度臨界点を求めることができる。

4 データセットとシステム

4.1 データセットと実装方法

防犯カメラの実映像データ [6] を用いた。このデータセットは、三つの店舗（電気店、コンビニ、銀行 ATM スタジオ）で撮影した映像からなる。各店舗では、三つのカメラによって撮影を行った。27 人が合計 56 の服装を着て現れており、人物ごとの服装数の最大は 9 パターンである。映像は合計で約 40 分である。本映像から、12180 枚のフレーム画像を抽出した。

次に、YOLO [13] を用いてフレーム画像から人体を検出し、DeepMAR [10] を拡張し人体の属性を推定した。本データセットにおいては 14926 の人体が検出され、各人体に対して 105 個の属性に関する確信度が推定された。属性の例として、年齢、性別、服の色、靴、アクセサリ、持ち物などが挙げられる。

ここで、推定された 105 個の属性の特徴量はバイナリ（Yes または No）である。よって、各属性に関する類似度は、問合せと特徴量が同一であれば 1、異なれば 0 とした。各属性の確信度において、閾値 0.5 で人物オブジェクトをクラスタリングし、式 (2) を用いて各属性の重みを計算した。

4.2 システム

図 3 はプロトタイプシステムの画面を示している。ユーザは画像または映像をアップロードして、属性推定の解析リクエストを出すと、システムは解析リクエストを受け付け、画像または映像から人物と属性を解析し保存する。

ユーザは三つのモードで検索できる：テキスト、画像、属性を選択。テキストで検索する場合、ユーザはあらかじめ決められたフォーマットで属性と確信度のテキストを入力する。これは、特定の属性条件を繰り返して利用する際に有用である。

画像で検索する場合、ユーザは対象人物を映した画像をアップロードして検索する。システムはアップロードされた画像から人物を検出し、その人物の属性と確信度を推定して属性条件として設定し、検索する。ユーザは推定された属性と確信度を手動で変更して検索することもできる。

また、ユーザは図 3 の左のパネルで、手動で属性を選択して確信度を入力して検索することもできる。図 3 では、プレセッ

トされた三つの属性のカテゴリが表示されており、高度な設定で他の属性を追加することもできる。あいまいな属性を検索条件に追加する場合は、確信度を下げることによってシステムに意図を反映できる。

ユーザからの問合せに対し、システムは問合せを属性と確信度のペアに変換する。検索処理では、まず候補の人物を抽出しマッチングスコアを計算する。ここで、問合せ属性に対し確信度スコアがゼロではない属性を有する人物が候補となる。その後、確信度や重みを考慮してマッチングスコアを計算し、スコアの高い順に人物候補をランキングする。最後に、上位の結果に対して各問合せ属性に関する確信度臨界点を計算し、検索結果と各問合せ属性の確信度臨界点を表示する。

図 3 の右では、検索上位の結果画像とそれらのマッチングスコアとフレーム時刻が示されている。各結果画像をマウスで指定すると、その人物画像から推定された属性が表示され、さらに問合せ属性と一致するものは強調して表示される。各結果画像の左下の一番目のアイコンをクリックすると、この画像から推定された属性と確信度を用いて検索するショートカットが提供されている。

連続したフレーム画像がデータベースに含まれる場合、検出された人物結果が同一人物であり重複することがあり、このままではユーザが結果を確認する手間が増えてしまう。重複を減らせるために、計算した結果または抽出した人体と属性に対して追跡を適用することができる。この問題に対し、本システムでは、Intersection over Union (IoU) に基づいた追跡アルゴリズム SORT [1] を適用し、連続したフレーム画像から検出された同一と思われる人物の結果をグループにまとめて表示することで対処した。なお、各結果画像の左下の二番目のアイコンをクリックするとグループされた結果画像を確認することができる。

さらに、各結果画像をクリックした場合、システムは結果人物が現れた時間前後の短い（5 秒程度）映像を表示する。

各問合せ属性に対して確信度臨界点が計算され検索画面に表示される。図 3 の左のパネルでは、グレーの矩形で各問合せ属性の確信度臨界点が表示されている。確信度臨界点の数が大きい場合（例えば、10）、ユーザに対してすべての確信度臨界点を提示することは好ましくない。そのため、本システムでは近い確信度臨界点をまとめて表示する。

ユーザは確信度のバーをスライドさせることで検索結果のランキング順位の変化を確認することができる。また、ある問合せ属性の確信度を調整した場合、他の問合せ属性の確信度臨界点も変わる場合がある。その際、システムは問合せ属性の確信度臨界点を再計算しユーザに提示する。

5 まとめと今後の課題

本論文では、問合せ属性のあいまい性と画像から推定された属性の不確実性を考慮した人物画像検索技術及びそれを実装したプロトタイプデモシステムについて紹介する。本研究は、実環境を想定した場合に発生する属性のあいまい性を有効に活用

Person Search by Uncertain Attribute

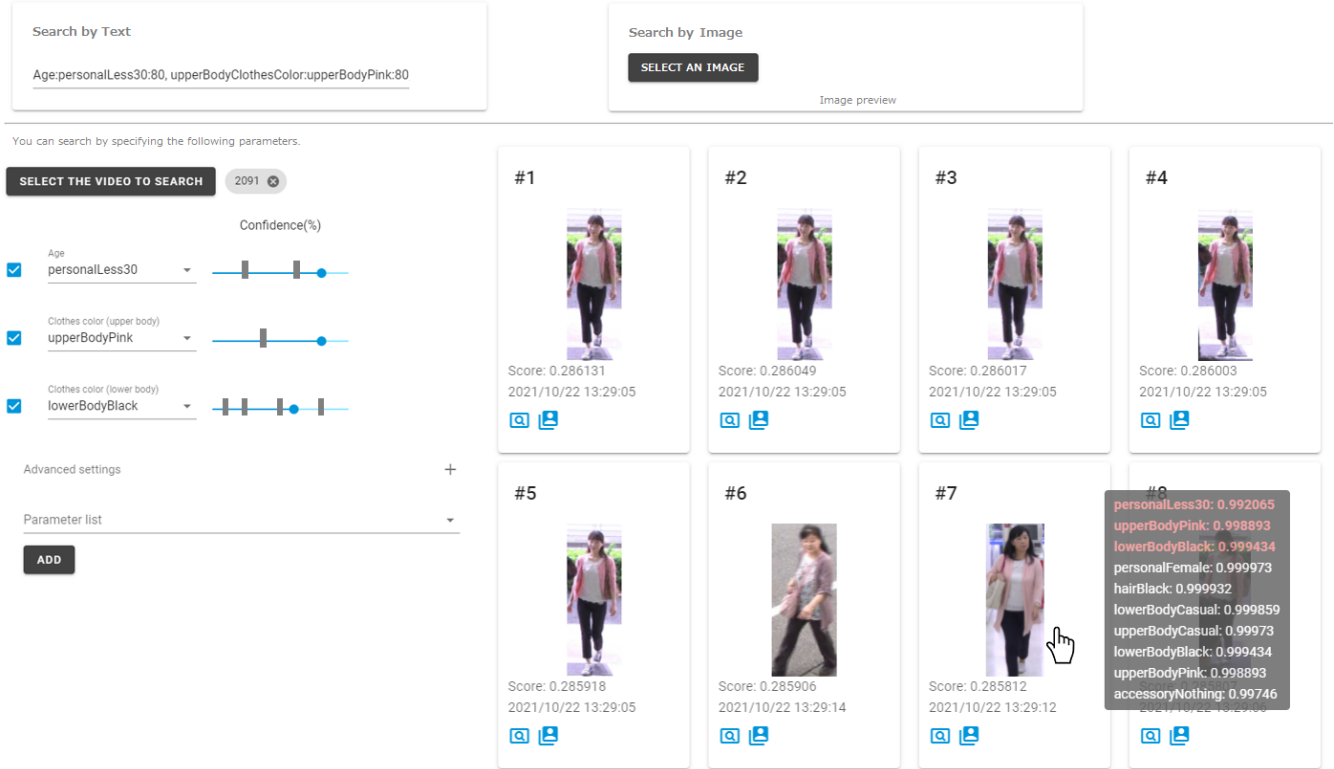


図3 プロトタイプシステムの画面. ユーザはテキスト, 画像, または属性条件を選択して入力して検索できる. 右のパネルにはマッチングスコアが高い順に検索結果が表示される. 問合せに用いた属性と, 計算された確信度臨界点は左のパネルに表示される. ある結果をマウスで指定すると, その画像から推定された属性が表示される. 問合せに用いた属性は強調して提示されている.

し人物画像検索の性能を高めることを目指す. 新たなマッチングスコアを定義し, ユーザが指定する問合せ属性のあいまい性, 及び属性推定技術により出力される属性の不確実性の両方を加味した検索を可能にした. また, 新たな定義として確信度臨界点を提案し, 問合せ属性の確信度が検索結果のランキング順位に与えるインパクトを可視化した. ウェブ上で動作するデモシステムを開発し, 実映像データを用いたテストを行い, 本手法の有用性を検証した. 今後は, 人物画像検索分野の公開データセットで評価実験を行い, 提案手法の有効性を証明する. さらに, 大規模な映像および複数の属性データに対する拡張性を向上し, 検索手法の効率性を改善する.

文献

- [1] Alex Bewley, Zongyuan Ge, Lionel Ott, Fabio Ramos, and Ben Upcroft. Simple online and realtime tracking. In *Proc. IEEE International Conference on Image Processing (ICIP)*, 2016.
- [2] Yu-Tong Cao, Jingya Wang, and Dacheng Tao. Symbiotic adversarial learning for attribute-based person search. In *Proc. European Conference on Computer Vision (ECCV)*, 2020.
- [3] Di Chen, Shanshan Zhang, Jian Yang, and Bernt Schiele. Norm-aware embedding for efficient person search. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [4] Mark de Berg, Otfried Cheong, Marc J. van Kreveld, and Mark H. Overmars. *Computational geometry: algorithms and applications*, 3rd Edition. Springer, 2008.
- [5] Qi Dong, Xiatian Zhu, and Shaogang Gong. Person search by text attribute query as zero-shot learning. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.
- [6] Tingting Dong, Shoji Nishimura, and Jianquan Liu. Refining image search results using multiple attributes. In *Proc. IEEE International Conference on Big Data (Big Data)*, 2019.
- [7] Xiao Han, Sen He, Li Zhang, and Tao Xiang. Text-based person search with limited data. *CoRR*, abs/2110.10807, 2021.
- [8] Boseung Jeong, Jicheol Park, and Suha Kwak. ASMR: Learning attribute-based person search with adaptive semantic margin regularizer. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.
- [9] Xu Lan, Xiatian Zhu, and Shaogang Gong. Person search by multi-scale matching. In *Proc. European Conference on Computer Vision (ECCV)*, 2018.
- [10] Dangwei Li, Xiaotang Chen, and Kaiqi Huang. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios. In *Proc. IEEE Asian Conference on Pattern Recognition (ACPR)*, 2015.
- [11] Shuang Li, Tong Xiao, Hongsheng Li, Bolei Zhou, Dayu Yue, and Xiaogang Wang. Person search with natural language description. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [12] Bharti Munjal, Sikandar Amin, Federico Tombari, and Fabio Galasso. Query-guided end-to-end person search. In *Proc. IEEE/CVF Conference on Computer Vision and*

- Pattern Recognition (CVPR)*, 2019.
- [13] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
 - [14] Dan Suciu. Probabilistic databases. In Ling Liu and M. Tamer Özsu, editors, *Encyclopedia of Database Systems, Second Edition*. Springer, 2018.
 - [15] Chufeng Tang, Lu Sheng, Zhaoxiang Zhang, and Xiaolin Hu. Improving pedestrian attribute recognition with weakly-supervised multi-scale attribute-specific localization. In *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.
 - [16] Yichao Yan, Jinpeng Li, Jie Qin, Song Bai, Shengcai Liao, Li Liu, Fan Zhu, and Ling Shao. Anchor-free person search. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.