

防犯カメラ映像における 条件分割型適合性フィードバックによる特定人物画像検索

井関 洋平^{1,a)} 川西 康友² 椋木 雅之² 美濃 導彦²

概要: 様々な環境に設置された防犯カメラで観測された人物画像の特徴量は被写体の姿勢、照明などの撮影条件によって大きく変化する。そのため、人物照合には撮影条件によって (1) 同一人物の人物画像の特徴量間の距離が大きい、(2) 別人同士の人物画像の特徴量間の距離が小さい、という 2 つの問題がある。特定人物画像検索において、従来提案されてきた適合性フィードバックでは、問題 (1) には対応できるが、問題 (2) には対処できない。本稿で提案する条件分割型適合性フィードバックでは、各防犯カメラでの人物追跡で得られる情報をもとに、撮影条件を分類してフィードバックすることで、各撮影条件の画像特徴量が混合されることを回避し、問題 (1)(2) の両方に対処する。複数の防犯カメラ映像に対して人物画像を検索し、通常の適合性フィードバックと比較することで本手法の有効性を確認した。

キーワード: 類似画像検索, 特定人物画像検索, 適合性フィードバック

1. はじめに

近年、防犯、犯罪行為の証拠確保、歩行者動向分析などを目的として、金融機関、空港、繁華街など街の至る所に、数多くの固定カメラ（防犯カメラ）が設置されている。これらの防犯カメラ映像は、犯罪者追跡や迷子捜索などの特定人物の足取り調査にも利用されている。たとえば、2012年5月21日に東京メトロ副都心線渋谷駅で起きた殺人未遂事件では、警察が120台以上の防犯カメラ映像を解析して、容疑者の逃走経路を割り出すことで、容疑者逮捕に至った。こうした調査には、できるだけ漏れなくその人物が映っている映像を見つけ出し、いつ、どのカメラでその人物が映ったのかを知る必要がある。従来、こうした調査は人手で行われてきたが、これには膨大な時間とコストが必要となる。

一方で、近年防犯カメラ映像を対象とした人物照合に関する研究 [1] や類似画像検索に関する研究 [2] が、コンピュータビジョンの分野で盛んに行われている。これらの技術を応用し、防犯カメラ映像中から特定の人物が映っているフレームを検索することで、防犯カメラ映像解析によ

る特定の人物の足取り調査を支援することが出来れば、人手による作業が大幅に減り、有用であると考えられる。

これを実現するためには、防犯カメラで観測された人物の人物画像（人物を外接矩形で切り出した画像）を登録した人物画像データベースを作成し、その中からある特定の人物の人物画像を全て検索すれば良い。本研究では、このように人物画像データベースを作成し、その中から特定人物の人物画像を全て検索することを特定人物画像検索と呼ぶ。

従来、特定人物画像検索の方法がいくつか提案されている [3-6]。これらの方法を一般化すると、ユーザ（検索を実行する人物）が用意したクエリ画像（1枚の被検索者の人物画像）とデータベース内の全ての人物画像の距離を人物照合を用いて計算し、距離が小さい順にその人物画像をユーザに対して提示するという手順であるといえる。

ここで、特定人物画像検索における人物照合には2つの問題点がある。1つ目の問題点は、同一人物の人物画像の特徴量同士でも撮影条件が異なれば、特徴量間の距離が離れてしまうことである。2つ目の問題点は、別人同士の人物画像の特徴量が、撮影条件によっては類似してしまう場合があることである。

この2つの問題点に対処する試みとして、Metternichら [5]、Fischerら [6] は特定人物画像検索において適合性フィードバックを用いて、検索結果に含まれる様々な撮影条件の下での被検索者の人物画像をクエリに反映すること

¹ 京都大学工学部情報学科
The School of Informatics and Mathematical Science, Faculty of Engineering, Kyoto University

² 京都大学学術情報メディアセンター
Academic Center for Computing and Media Studies, Kyoto University

a) iseki@mm.media.kyoto-u.ac.jp

で、人物照合の1つ目の問題点に対処することを提案している。しかし、適合性フィードバックは2つ目の問題点には対処することができない。

そこで、本研究では、特定人物画像検索における人物照合の1つ目の問題点に対処できる「適合性フィードバック」と、人物照合を撮影条件ごとに分割して行うことで、2つ目の問題点に対処できる「条件分割による検索」を組み合わせた条件分割型適合性フィードバックを提案し、特定人物画像検索における人物照合の両方の問題点に対処する。

2. 特定人物画像検索

2.1 特定人物画像検索の一般的な手順

従来、特定人物画像検索の方法はいくつか提案されている [3-6]。これらを一般化した特定人物画像検索を実現するシステムの処理の手順を図 1 に示す。この手順は、防犯カメラ映像から人物画像データベースを作成する「データベース作成段階」と、ユーザがクエリ画像を用意して検索を実行し、結果を得る「検索実行段階」に分かれる。

防犯カメラ映像の各フレームに対して、人物検出・人物追跡を用いることで、ある人物がカメラに映り始めてから映り終わるまでのトラッキング系列の人物画像列が得られる。特定人物画像検索のデータベース作成段階では、システムがこれを全て人物・防犯カメラ映像に対して行うことで、トラッキング系列の人物画像列の集合 $\{I_1, \dots, I_n\} (I_k = \{i_{k1}, \dots, i_{kn_k}\})$ を得る。ここで、 n は人物画像列の数、 I_k は k 番目の人物画像列、 i_{kl} は人物画像列 I_k の l 番目の人物画像、 n_k は人物画像列 I_k に含まれる人物画像の数を表す。次に、各人物画像列 I_k の全ての人物画像に対して特徴量抽出を行い、人物画像列 I_k に対応する特徴量列 $F_k = \{f_{k1}, \dots, f_{kn_k}\}$ (ただし f_{kl} は人物画像 i_{kl} の特徴量) を得る。そして、人物画像列 I_k 、それに対応する特徴量列 F_k 、人物画像列 I_k が観測されたカメラ ID x_{ck} 、時刻情報 x_{tk} で構成されるレコード R_k を作成し、それらをデータベースに登録する。このとき、データベースには R_1, \dots, R_n の n 個のレコードが登録される。

検索実行段階では、ユーザがクエリ画像 i_q を用意して、検索を実行する。システムはクエリ画像 i_q からクエリ特徴量 f_q を抽出する。そして、データベース内の各レコード R_k を、人物照合を用いて計算できるクエリ特徴量-レコード間距離 $D(f_q, R_k)$ が小さい順に検索結果としてユーザに提示する。レコードをユーザに提示する際は、レコード内に含まれる全ての人物画像を提示する。ここで、クエリ特徴量-レコード間距離 $D(f_q, R_k)$ は、クエリ特徴量とそのレコード内に含まれる特徴量列との距離 $d(f_q, F_k)$ と定義される。

2.2 特定人物画像検索における人物照合の問題点

特定人物画像検索において、クエリ特徴量-レコード間距

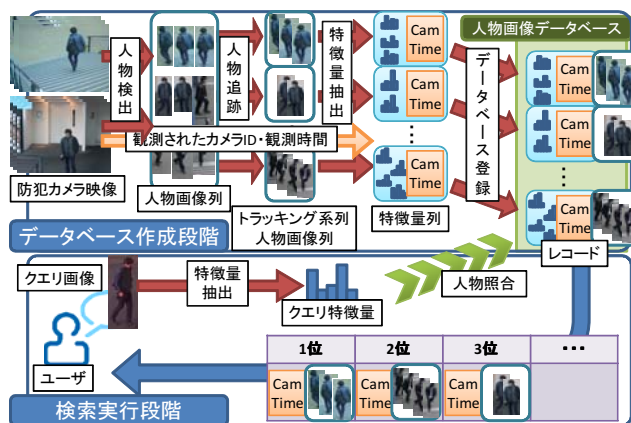


図 1: 一般的な特定人物画像検索の手順

離 $D(f_q, R_k)$ の計算には、人物照合による人物画像の特徴量比較が必要であるが、一般に人物照合における特徴量比較は難しい問題とされている。たとえば、人物照合の評価で用いられるデータセットの一つ、VIPeR [7] を用いた人物照合率は、最新の手法を用いても、20%程度であることが知られている。これは、人物の特徴量が照明、遮蔽の有無、被写体の姿勢、解像度などの撮影条件によって大きく変化する事により、人物照合の際に次の2つの問題が生じるためである。1つ目の問題は、同一人物の人物画像の特徴量同士でも撮影条件が異なれば、特徴量間の距離が離れてしまうことである。このとき、被検索者の人物画像を含むレコードのうち、クエリ画像の撮影条件と類似した撮影条件の人物画像を含むレコードしか検索上位にならない。2つ目の問題は、本来違う特徴量を持つ別人の人物画像の特徴量が、撮影条件によっては被検索者の人物画像の特徴量と類似してしまう場合があることである。このとき、そうした本来違う特徴量を持つ別人の人物画像を含むレコードが、検索上位になってしまう。

3. 特定人物画像検索における従来手法

Metternich ら [5] や Fischer ら [6] は特定人物画像検索に対して適合性フィードバックを用いて、検索精度を向上させることを提案している。適合性フィードバックは、ユーザに対して検索結果の適合/不適合のフィードバックを求め、それを検索時の順位付けに反映する手法である。適合性フィードバックを用いることで、検索時にユーザの助けを借りて、高い精度で検索を行うことができる。

Metternich ら [5] や Fischer ら [6] の提案する方法を一般化した、適合性フィードバックを用いた特定人物画像検索を実現するシステムの処理の手順は以下の通りである (図 2)。検索開始時をステップ $t = 0$ とし、その時のクエリ特徴量 $f_q(t)$ は、ユーザが用意したクエリ画像 i_q から抽出した特徴量とする。

(1) システムは、現在のクエリ特徴量 $f_q(t)$ を用いてデータベース内のユーザにまだ提示していないすべてのレ

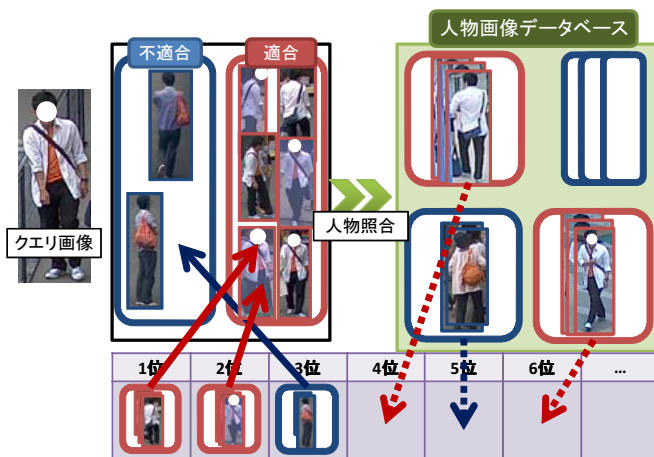


図 2: 一般的な適合性フィードバックを用いた特定人物画像検索を実現するシステムの手順 ($N = 3$ の場合)

コードについて、それらのレコードをクエリ特徴量-レコード間距離 $D(\mathbf{f}_q, R_k)$ の小さい順に N 件を検索結果として、ユーザに提示する。

- (2) ユーザは提示された検索結果に対して適合／不適合のフィードバックを行う。このとき、レコード単位ではなくレコード内の人物画像単位でフィードバックを行う。システムは、検索結果の各レコード内に含まれる全ての特徴量を、ユーザのフィードバックをもとに、適合集合 S^+ 、不適合集合 S^- に振り分ける。
- (3) システムは適合集合 S^+ 、不適合集合 S^- に含まれる特徴量をクエリ特徴量に反映し、クエリ特徴量を更新する。ここで、適合集合 S^+ 、不適合集合 S^- に含まれる特徴量をクエリ特徴量に反映する方法は、適合性フィードバックの手法によって異なる。そして、 $t \leftarrow t+1$ として、(1) に戻る。

これをユーザが検索をやめるまで繰り返す。

Metternich ら [5] は適合性フィードバック手法の中で、クエリ特徴量を更新する手法として、クエリ最適化が特定人物画像検索において有効であるとしている。クエリ最適化は、以下の Rocchio の式 [8] に基づいて、1 ステップごとにクエリ特徴量を更新する。

$$\mathbf{f}_q(t+1) = \alpha \mathbf{f}_q(t) + \frac{\beta}{|S^+|} \left(\sum_{\mathbf{s}_i^+ \in S^+} \mathbf{s}_i^+ \right) - \frac{\gamma}{|S^-|} \left(\sum_{\mathbf{s}_i^- \in S^-} \mathbf{s}_i^- \right) \quad (1)$$

ここで、 α, β, γ はクエリ最適化の際のフィードバックの重みを表す。クエリ特徴量に対して、ユーザが適合とした人物画像の特徴量の平均を加算し、不適合とした人物画像の特徴量の平均を減算することで、クエリ特徴量を更新する。

データベース内の各レコードには、様々な撮影条件の人物画像が含まれている。たとえば、ある人物がカメラ内で別方向を向けば、被写体の方向が異なる人物画像が一つのレコードに含まれる。また、その人物が日陰から日向に移

動すれば、照明環境が異なる人物画像が一つのレコードに含まれる。

そのため、クエリ最適化によって適合性フィードバックを用いて検索することで、クエリ画像の撮影条件とは異なる撮影条件下で撮影された被検索者の人物画像の特徴量をクエリ特徴量に反映させることできる。このとき、クエリ特徴量は被検索者の様々な撮影条件の下での人物画像の特徴量の平均に近くなる。そのため、クエリ特徴量と被検索者の様々な撮影条件の下での人物画像の特徴量との距離の平均が小さくなる。よって、異なる撮影条件の同一人物の人物画像の特徴量間の距離が大きくなるという特定人物画像検索における人物照合の1つ目の問題に対処できる。

しかし、適合性フィードバックはユーザがフィードバックした人物画像の特徴量のみ注目しているため、本来違う特徴量を持つ別人の人物画像の特徴量が、撮影条件によっては被検索者の人物画像の特徴量と類似してしまう場合があるという人物照合の2つ目の問題には対処できない。そのため、被検索者の人物画像が含まれるレコードを多く得るためには、それだけ多くの検索結果を見る必要があり、検索精度が低い。

4. 条件分割型適合性フィードバック

本稿では、検索結果に含まれる様々な撮影条件の下での被検索者の人物画像の特徴量をクエリ特徴量に反映することで、特定人物画像検索における人物照合の1つ目の問題点に対処できる「適合性フィードバック」と、人物照合を撮影条件ごとに分割して行うことで2つ目の問題点に対処できる「条件分割による検索」を組み合わせた条件分割型適合性フィードバックを提案し、特定人物検索における人物照合の問題点の両方に対処する。

4.1 条件分割による検索

人物照合の2つ目の問題は、本来違う特徴量を持つ別人の人物画像の特徴量が、撮影条件によっては被検索者の人物画像の特徴量と類似してしまう場合があることである。そのため、人物照合を撮影条件ごとに分割して行うことで、ある撮影条件の下で撮影された被検索者の人物画像の特徴量分布と、別の撮影条件の下で撮影された別人の人物画像の特徴量分布が重なっていたとしても、重なった分布に存在する別人の人物画像を含むレコードが検索上位に現れることを抑えられ、検索精度が向上すると考えられる。例えば、図3(a)のように、クエリ画像の人物の明るいシーンで観測された人物画像と、暗いシーンで撮影された別人の人物画像(図3(a)右下のレコードに含まれる人物画像)が類似している時、この人物の人物画像が含まれるレコードが検索上位になってしまう。しかし、同じ暗いシーン同士の人物画像を比べると、特徴量が類似していないことが分かる。

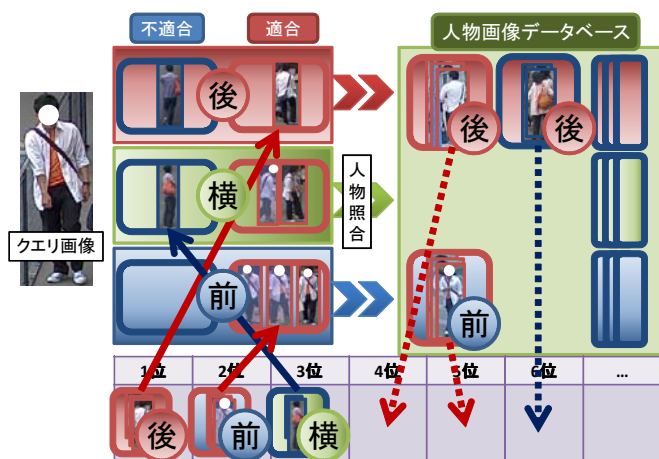


図 4: 条件分割型適合性フィードバックを用いた検索

このとき、図 3(b) のように、クエリ特徴量を照明条件の明・暗で分割し、明るい撮影条件の人物画像は明るい撮影条件のクエリ特徴量と、暗い撮影条件の人物画像は暗い撮影条件のクエリ特徴量を用いて人物照合を行うことで、図 3(b) の右下のレコードが検索上位に現れなくなり、検索精度が向上すると考えられる。

この条件分割による検索を実現するためには、事前にデータベース内の人物画像に対して、撮影条件を付与する必要がある。ここで、一般的に特徴量そのものからその画像の撮影条件を知ることは難しい。しかし、特定人物画像検索においては、データベース作成段階で人物追跡を行うことで、レコード内に含まれる人物画像の防犯カメラ映像中の座標や、撮影された時刻、同レコード内の別画像に対する特徴量の相対的關係など、特徴量そのもの以外の情報も得られる。そこで、予め定めた m 個の撮影条件 $C = \{c_1, \dots, c_m\}$ について、人物追跡で得られる情報をもとに、データベース内の各レコード R_k に含まれる各人物画像とそれに対応する特徴量の組 $(i_{kl}, \mathbf{f}_{kl})$ に対して、撮影条件 $c_{kl} \in C$ を付与する。この時得られる各レコード R_k の人物画像列 I_k 、特徴量列 F_k に対応する撮影条件列 $C_k = \{c_{k1}, \dots, c_{kn_k}\}$ を、レコード R_k に追加する。

また、条件分割による検索には、条件分割に用いる各撮影条件 c_i の下での被検索者の人物画像を用意する必要がある。これは、犯罪者追跡や迷子捜索などの特定人物の足取り調査を行う際には、大きな手間となる。

4.2 条件分割型適合性フィードバックによる検索

そこで、本研究では、「適合性フィードバック」と「条件分割による検索」を組み合わせ、条件分割に用いる撮影条件の下での人物画像を適合性フィードバックを用いて獲得することで、人物照合の両方の問題に対応した条件分割型適合性フィードバック (Condition-Separating Relevance Feedback: CSRF) を提案する (図 4)。

CSRF を用いて特定人物画像検索を行うシステムは、ま

ず 4.1 節で述べたように、データベース作成段階で、予め定めた m 個の撮影条件 $C = \{c_1, \dots, c_m\}$ について、人物追跡で得られる情報を基に、データベース内の各レコード内 R_k に、撮影条件列 $C_k = \{c_{k1}, \dots, c_{kn_k}\} (c_{kl} \in C)$ を追加する。

次に、検索実行段階で、初めにユーザが用意したクエリ画像 i_q から抽出したクエリ特徴量 \mathbf{f}_q をもとに、先に定めた各撮影条件 c_1, \dots, c_m に対応する m 個のクエリ特徴量 $\mathbf{f}_{qc_1}(t), \dots, \mathbf{f}_{qc_m}(t)$ を複製して作成する (ここで、 $\mathbf{f}_{qc_i}(0) = \mathbf{f}_q$)。CSRF を用いた検索では、これらのクエリ特徴量をまとめた、クエリ特徴量集合 $Q(t) = \{\mathbf{f}_{qc_1}(t), \dots, \mathbf{f}_{qc_m}(t)\}$ を用いて検索を行う。

CSRF を用いた検索では、適合性フィードバックを用いた検索と同様に、ユーザが検索結果に対して適合／不適合のフィードバックを行う。このとき、システムは検索結果の各レコード内に含まれる全ての特徴量を、ユーザのフィードバックとその特徴量に対応する人物画像の撮影条件をもとに、 m 個の撮影条件に対応する適合集合 $S_{c_1}^+, \dots, S_{c_m}^+$ 、不適合集合 $S_{c_1}^-, \dots, S_{c_m}^-$ に振り分ける。たとえば、システムが検索結果として提示したレコード R_k の l 番目の人物画像 i_{kl} をユーザが適合と判断した場合、システムは、その特徴量 \mathbf{f}_{kl} を、その撮影条件 c_{kl} に対応する適合集合 $S_{c_{kl}}^+$ に加える。

そして、各撮影条件に対応する適合集合 $S_{c_i}^+$ 、不適合集合 $S_{c_i}^-$ に含まれる特徴量を後述の式 (4) を用いて、各クエリ特徴量 $\mathbf{f}_{qc_i}(t)$ に反映させることで、対応する撮影条件に特化した各クエリ特徴量 $\mathbf{f}_{qc_i}(t+1)$ に更新する。さらに、クエリ特徴量集合 $Q(t)$ とレコード R_k 内の特徴量列 F_k との距離計算には、各特徴量 \mathbf{f}_{kl} と同じ撮影条件 c_{kl} に対応するクエリ特徴量 $\mathbf{f}_{qc_{kl}}$ を用いて比較することで、ある撮影条件の下で撮影された被検索者の人物画像の特徴量分布と、他の撮影条件の下で撮影された別人の人物画像の特徴量分布が重なっていたとしても、重なった分布に存在する別人の人物画像を持つレコードが検索上位に来ることを防ぎ、検索精度の向上を図る。このとき、クエリ特徴量集合 $Q(t)$ と各レコード R_k との距離 $D(Q(t), R_k)$ は以下の式で計算する。

$$D(Q(t), R_k) = \min_{c_i \in C} d(\mathbf{f}_{qc_i}(t), F_{kc_i}) \quad (2)$$

ここで、 $F_{kc_i} = \{\mathbf{f}_{kj} | \forall j, \mathbf{f}_{kj} \in F_k, c_{kj} \in C_k, c_{kj} = c_i\}$ 。

4.2.1 フィードバック方法の改善

「適合性フィードバック」と「条件分割による検索」を組み合わせる時、クエリ特徴量を更新する際に単純に式 (1) を条件分割に対応させた式 (3) を用いて、各撮影条件に対応するクエリ特徴量 $\mathbf{f}_{qc_i}(t)$ に対して、同じ撮影条件 c_i に対応する適合集合 $S_{c_i}^+$ 、不適合集合 $S_{c_i}^-$ に属する特徴量のみを反映することが考えられる。

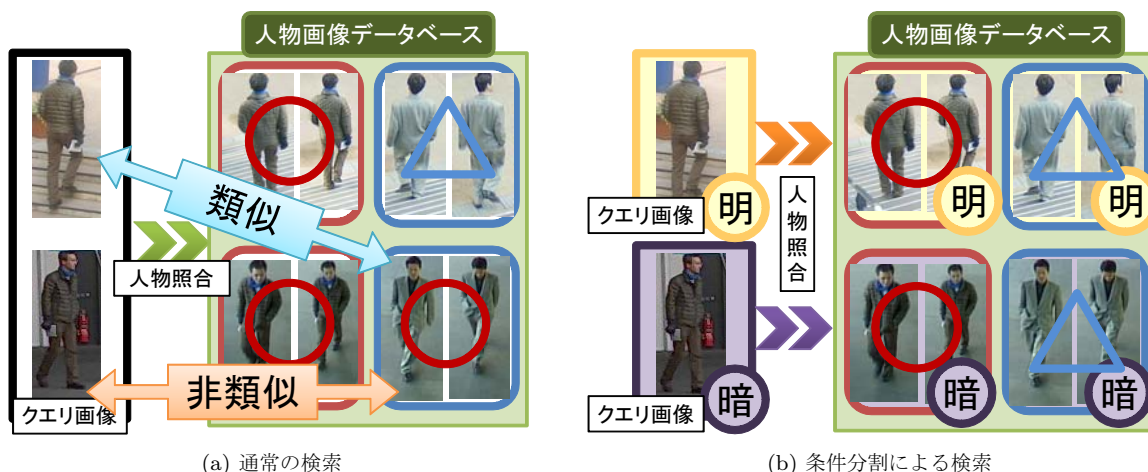


図 3: 条件分割による人物照合の 2 つ目の問題点への対処

$$\mathbf{f}_{q_{c_i}}(t+1) = \alpha \mathbf{f}_{q_{c_i}}(t) + \frac{\beta}{|S_{c_i}^+|} \left(\sum_{s_{c_i}^+ \in S_{c_i}^+} s_{c_i}^+ \right) - \frac{\gamma}{|S_{c_i}^-|} \left(\sum_{s_{c_i}^- \in S_{c_i}^-} s_{c_i}^- \right) \quad (3)$$

しかし、この式を用いてクエリ特徴量を更新すると、各クエリ特徴量 $\mathbf{f}_{q_{c_i}}(t)$ へのフィードバック件数 $|S_{c_i}^+| + |S_{c_i}^-|$ が各撮影条件ごとに分散してしまうという問題が生じる。

CSRF を用いた検索では、各撮影条件に対応するクエリ特徴量 $\mathbf{f}_{q_{c_i}}(t)$ はフィードバックされた撮影条件 c_i の人物画像の特徴量の平均に近づく。そのため、フィードバック件数が多い場合は、撮影条件 c_i の下での被検索者の人物画像の特徴量分布の中心に近くなる。

しかし、フィードバック件数が少ない場合は、その効果が薄れてしまうため、フィードバック件数が少ない撮影条件に対応するクエリ特徴量に対しては、類似した撮影条件へのフィードバックも、そのクエリ特徴量に反映した方が、その撮影条件の下での被検索者の特徴量分布の中心に近くなる可能性がある。

そこで、CSRF を用いた検索では、各撮影条件に対応するクエリ特徴量 $\mathbf{f}_{q_{c_i}}(t)$ に対して、以下の式を用いてフィードバックを行うことで、この問題に対処する。

$$\mathbf{f}_{q_{c_j}}(t+1) = \alpha \mathbf{f}_{q_{c_j}}(t) + \sum_i \omega_{ij} \left\{ \frac{\beta}{|S_{c_i}^+|} \left(\sum_{s_{c_i}^+ \in S_{c_i}^+} s_{c_i}^+ \right) - \frac{\gamma}{|S_{c_i}^-|} \left(\sum_{s_{c_i}^- \in S_{c_i}^-} s_{c_i}^- \right) \right\} \quad (4)$$

ここで、 α, β, γ はクエリ最適化の際のフィードバックの重み、 ω_{ij} は、撮影条件 c_i から撮影条件 c_j へのフィードバックの重みである。この式は、撮影条件 c_i に関するフィードバックを ω_{ij} という割合で重みづけして撮影条件 c_j に対応するクエリ特徴量 $\mathbf{f}_{q_{c_j}}(t+1)$ にも伝播させることを意味している。この式を用いることで、撮影条件間でフィードバック件数に偏りがある場合にも、他の撮影条件からの

フィードバックを伝播させて、各撮影条件に対応するクエリ特徴量を更新することができる。

5. 実験

防犯カメラ映像からの特定人物画像検索において、CSRF を用いることの有効性を確認するために、実際の商業施設に設置されている防犯カメラ映像に対して、特定人物画像検索を行った。

5.1 特定人物画像検索で用いた手法

5.1.1 データベース段階で用いた手法

特定人物画像検索では、多数の防犯カメラ映像を扱う。このとき、防犯カメラ映像が録画される時間よりも、その防犯カメラ映像を処理して、人物画像データベースを作成する時間の方が短くなければならない。一般に防犯カメラはほぼ毎日、12 時間以上稼働しているものも多い。また、1 台のコンピュータで複数の防犯カメラの映像を処理することも考えられる。そのため、データベース作成段階で用いる人物検出・人物追跡・特徴量抽出手法には、精度が比較的良く、その中で処理時間が短いものを選んだ。

5.1.1.1 人物検出・人物追跡

人物検出・人物追跡には高速に人物を検出・追跡できる OKAO Vision [9] の人体トラッキングソフトウェアを用いた。

5.1.1.2 特徴量・特徴量比較方法

人物画像の特徴量には、人物照合の分野で、実装が容易で精度が良いことで知られる Farenzena ら [10] の重み付き色ヒストグラムを用いた。HSV 色ヒストグラムのビン数は経験的に 3 チャンネルとも 32 とし、人物の特徴量を 32 ビン \times 3 チャンネル \times 2 領域の 192 次元の特徴量で表した。また、特徴量の比較には、Farenzena らが用いた Bhattacharyya 距離を用いた。

この手法にはシルエット抽出が必要であるが、データ



(a) 人物画像以外の画像 (b) 人物の外接矩形がずれている画像 (c) 遮蔽のある画像



(d) 二人以上の人物が映っている画像 (e) 別人の人物画像の組

図 5: 人物検出誤り (a)(b)(c)(d), 人物追跡誤り (e)

ベース作成段階で行われる処理は、計算時間が短いものが良い。そこで、本研究ではシルエット抽出に、画像全体の画素を使わずに背景画像を推定することで、ロバストかつ高速な背景差分が可能である GROUSE [11] を用いた。各パラメータとして、低次元空間の次元数を 2、ランダムな画素選択の際に全画素中から選ばれる画素の割合を 30% とした。

5.1.2 検索実行段階で用いた手法

どんな人物検出、人物追跡の手法を用いても 100% ミスのない検出、追跡をすることは難しい。そのため、防犯カメラ映像から人物検出・人物追跡を行って得た一つのトラッキング系列の人物画像列には、必ずしもある一人の人物の人物画像だけが含まれるとは限らない。人物検出によって得られたある画像が、「人物画像以外の画像」(図 5(a)), 「人物の外接矩形がずれている画像」(図 5(b)), 「遮蔽のある画像」(図 5(c)), 「二人以上の人物が映っている画像」(図 5(d)) であった場合を「人物検出誤り」と呼ぶ。また、あるレコードが別人の人物画像の組 (図 5(e)) を含む場合を「人物追跡誤り」と呼ぶ。そして、人物検出誤りを生じた画像を含んだレコード、人物追跡誤りを生じたレコードを総じて「誤りを含むレコード」と呼ぶ。

特定人物画像検索の検索実行段階で行うクエリ特徴量-レコード間距離比較は、誤りを含むレコードが人物画像データベースに存在しても、検索に影響を受けにくい手法を用いるのが望ましいと考えられる。

5.1.2.1 クエリ特徴量-レコード間距離

特定人物画像検索においてクエリ特徴量-レコード間距離は、1 対多の特徴量間の距離 $d(\mathbf{f}, F)$ を用いて求められる。ここで、 \mathbf{f} は一つの特徴量、 F は特徴量列 ($F = \mathbf{f}_1, \dots, \mathbf{f}_n$) である。本研究では、距離 $d(\mathbf{f}, F)$ を F に含まれる全ての特徴量と特徴量 \mathbf{f} との距離のうち最小のものをと定義する。

$$d(\mathbf{f}, F) = \min_k d(\mathbf{f}, \mathbf{f}_k) \quad (5)$$

この方法を用いることで、ある被検索者の人物画像が含まれるレコードが誤りを含むレコードであっても、そのレコード内に 1 枚でも被検索者の人物画像が含まれていれば、そのレコードを検索上位にすることができる。

5.2 条件分割に用いた撮影条件

CSRF で使用する条件は、その条件自体が何らかの方法で自動的に判別可能であることが必要である。また、その条件は人物画像の特徴量に対して大きく影響を与えることが望ましい。そこで、今回は防犯カメラ映像の各フレームに対して人物追跡を行うことで得られる情報を用いて判別可能で、人物画像の特徴量に影響を与える条件として、被写体の方向条件、照明条件の 2 つの条件を用いた。

5.2.1 被写体の方向による条件分割

被写体の方向条件は、人物画像に映っている人物が、半そでの服を着ている場合や背中にリュックサックを背負っている場合など、被写体の体の向きによって色や模様に変化する服装をしている場合に、人物画像の特徴量に影響を与えると考えられる。データベース内の各人物画像に対する被写体の方向条件の付与については、人物が移動する際には、体の向きが常に進行方向を向いているという仮定をもとに、防犯カメラ映像中の検出座標の移動方向によってその人物の向いている方向を判断する (表 1)。

被写体の方向条件は経験的に、前方向、横方向、後方向の 3 つに分割し、同条件へのフィードバックは 1.0 の重みを、他条件へのフィードバックは 0.6 の重みをかけることとした。

5.2.2 照明による条件分割

照明条件は、防犯カメラが観測しているシーンの照明の度合いを表し、人物画像の特徴量の中で主に色特徴に大きな影響を与える。例えば、灰色の服を着た人物の人物画像の特徴量は、照明条件によって黒色や白色の服を着た人物の人物画像の特徴量と類似した特徴量に変化する。光源を白色と仮定すると、照明条件による人物の特徴量の変化は、特徴量の中で輝度、彩度への影響が大きい。そこで、本研究では人物画像の照明条件を、その人物画像に映っている人物のすべての撮影条件の下での人物画像集合を考え、その中での人物画像の平均輝度、平均彩度の相対的な大小と考える。データベース内の各人物画像に対する照明条件の付与については、その人物のすべての撮影条件の下での人物画像集合を知ることができないため、各レコード内での人物画像の平均輝度、平均彩度の相対的な大小とする。あるレコード R_k の各人物画像 i_{kl} について、人物領域 (実際に人物が映っている領域) の平均輝度の大きさ v_{kl} 、平均彩度の大きさ s_{kl} を求め、レコード R_k 内の人物画像列 I_k の平均輝度列 $V_k = \{v_{k1}, \dots, v_{kn_k}\}$ 、平均彩度列

表 1: 被写体の方向条件の例

画像			
被写体の方向	前向き	横向き	後向き

$S_k = \{s_{k1}, \dots, s_{knk}\}$ を得る. そして, 平均輝度列 V_k , 平均彩度列 S_k をそれぞれ k -means 法で n_v, n_s 個にクラスタリングし, 各人物画像 i_{kl} の平均輝度 v_{kl} , 平均彩度 s_{kl} が属するクラスタの中心と, 他のクラスタ中心との大小に応じて, 各人物画像 i_{kl} のレコード R_k 内での相対的な輝度, 彩度の大小を判断する (表 2).

照明条件は経験的に, 輝度, 彩度ともに 3 つに分割し ($n_v = n_s = 3$), $3 \times 3 = 9$ 条件で, 照明条件を表現する. また, 輝度・彩度ともに一つ隣の条件に対して 0.7, 2 つ隣の条件に対して 0.4 の重みをかけることとした. 例えば, 輝度, 彩度がそれぞれ一つずつ違う条件のクエリ特徴量に対するフィードバックの重みは, $0.7 \times 0.7 = 0.49$ とした.

また, 検索結果の特徴量を各撮影条件に対応する適合集合 S_c^+ , 不適合集合 S_c^- に振り分ける際, 適合集合における照明条件に関しては, データベース作成段階で付与された撮影条件を用いず, 再度適合集合内でクラスタリングすることにより各特徴量の撮影条件を付け直し, 対応する各適合集合 S_c^+ に振り分けた. 不適合集合に関しては, データベース作成段階で付与された撮影条件をそのまま用いた.

5.3 実験に用いる人物画像データベース

本実験に用いた人物画像データベースは, 商業施設に設置された 22 台の防犯カメラで観測された 3 時間分の映像, 合計 66 時間分の映像に対して, 5.1 節で述べた人物検出・人物追跡・特徴量抽出を用いて作成した人物画像データベースである. 人物画像データベース内には, レコードが合計 $n=12673$ 個存在し, 総人物画像枚数は 368356 枚である. 人物画像データベース内には, 誤りを含むレコードも多数存在している.

5.4 検索対象

本実験で検索対象とする人物は, その人物の人物画像を含むレコードが人物画像データベース内に存在する特定の 10 人とした. 一人当たりのレコード数は 33~72 件で, 10 人の合計レコード数は 586 件である.

表 2: 照明条件の例

画像				
輝度	大	大	小	小
彩度	大	小	大	小

5.5 検索方法

まず, 人物画像データベース内の 10 人の検索対象となる人物の人物画像を含むレコードの中から, 異なるカメラから作成されたレコードを一人当たり 13~15 個, 合計 146 個選ぶ. 次に, 各レコード内に含まれる検索対象の人物の人物画像を一枚選び, 146 枚の人物画像をクエリ画像として抽出し, 146 回の検索を行う. このとき, クエリ画像を抽出したレコードは検索対象としない. ここで, クエリ画像に用いる画像は, 人物検出誤りを起こしていない人物画像を選んだ.

検索結果のフィードバックの際には, ユーザは被検索者の人物画像を確実に判断できると仮定し, 検索結果としてユーザに提示されたレコード内の人物画像の中に, 被検索者の人物画像が含まれていて, 人物検出誤りを起こしていない場合は, それを適合人物画像としてフィードバックした. ただし, レコード内に被検索者の人物画像が含まれているが, そのすべてが人物検出誤りを起こしている場合は, その中で最も人物検出誤りの度合いが最も小さい 1 枚を適合人物画像としてフィードバックした. ここで, 各検索において検索結果としてユーザに提示されたレコードに存在する人物画像が, 適合人物画像かどうかの判断は全て人手で行った.

また, 各パラメータとして, 適合性フィードバック 1 ステップごとにユーザに提示する検索結果レコード数 N の値は 20 とした. また, クエリ最適化のパラメータ α, β, γ の割合は経験的に $2:5:4$ ($\alpha + \beta + \gamma = 1$) とした. 全てのパラメータは通常の実験性フィードバック, CSRF のどちらも同じパラメータとした.

5.6 評価方法

3 章で述べた従来手法の通常の実験性最適化と, クエリ最適化に CSRF を用いた場合とで検索精度を比較した.

本研究が想定する特定の人物の足取り調査ではデータベース内にある被検索者の人物画像が含まれるレコードをできるだけ漏れなく見つけることが重要である. そこで, 評価方法は Metternich ら [5] と同様に, 検索順位に対する再現率を用いて評価する. 再現率は, データベース内に含

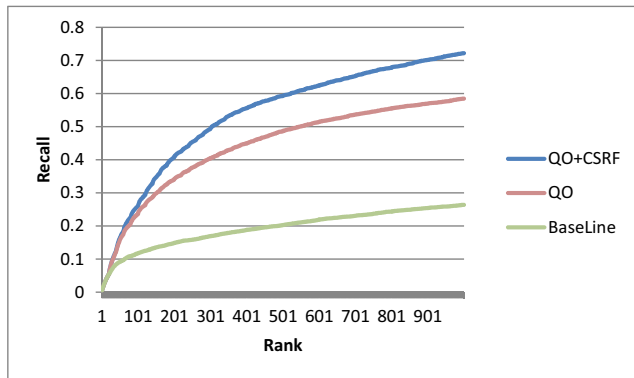


図 6: 実験結果

まれる全適合レコード数のうち、何レコードが見つかったのかの割合を示す。ここで、適合レコードは適合人物画像が一枚以上含まれるレコードのことを指す。

146 回の検索結果について、検索対象となる人物ごとに検索結果を平均し、さらにそれを 10 人で平均したものを、グラフにプロットして比較した。

5.7 実験結果

通常のクエリ最適化を用いた特定人物画像検索とクエリ最適化に CSRF を用いた特定人物画像検索の比較を図 6 に示す。BaseLine は適合性フィードバックを使用しない検索、QO(QueryOptimization) は通常のクエリ最適化による検索、QO+CSRF は、クエリ最適化に被写体の方向、照明条件で CSRF を使用した場合の結果を示す。

実験の結果、クエリ最適化による適合性フィードバックを用いて検索した方が、BaseLine よりも、上位 1000 件目での再現率が 26.4%から、58.5%へ 32.1 ポイント上昇した。また、クエリ最適化に被写体の方向条件、照明条件で CSRF を用いて検索を行うことによって、従来のクエリ最適化よりも、上位 1000 件目での再現率が 58.5%から、72.2%へ 13.7 ポイント上昇した。よって、防犯カメラ映像からの特定人物画像検索において、CSRF が有効であることが確認できた。

6. まとめ

特定人物画像検索における人物照合には、人物画像の特徴量が撮影条件により変化するために、(1) 同一人物の人物画像の特徴量間の距離が大きい (2) 別人同士の人物画像の特徴量間の距離が小さい、という二つの問題が存在する。従来提案されてきた適合性フィードバックでは、問題点 (1) に対処できるが、問題点 (2) に対処することはできない。そこで本研究では、検索結果に含まれる様々な撮影条件の下での被検索者の人物画像をクエリに反映することで、問題点 (1) に対応する「適合性フィードバック」と人物照合を撮影条件ごとに分割して行うことで問題点 (2) に対処できる「条件分割による検索」を組み合わせた条件分

割型適合性フィードバックを提案し、両方の問題に対処した。そして、実際の商業施設に設置された防犯カメラ映像に対して特定人物画像検索を行い、従来手法である通常の適合性フィードバックと比較することで、その有効性を示した。

今後の課題として、特定人物画像検索において人物画像の特徴量として服装の見え特徴だけでなく、顔特徴や歩容特徴などの複数の特徴量を用いることでの検索精度向上や、SVM などの学習手法を用いた適合性フィードバックの手法での、条件分割型適合性フィードバックの適用が考えられる。

謝辞 本研究は文部科学省、安心・安全な社会のための犯罪・テロ対策技術等を実用化するプログラム「環境適応型で実用的な人物照合問題」の一環として実施したものである。

参考文献

- [1] 井尻善久, 川西康友, 美濃導彦, 村瀬洋: サーベイ論文: 視野を共有しない複数カメラ間での人物照合, 電子情報通信学会パターン認識・メディア研究会 (PRMU), Vol. 111, No. 317, PRMU 2011-119, pp. 117-124 (2011).
- [2] Schettini, R., Ciocca, G., Zuffi, S. et al.: A survey of methods for colour image indexing and retrieval in image databases, *Color Imaging Science: Exploiting Digital Media*, pp. 183-211 (2001).
- [3] Bazzani, L., Cristani, M., Perina, A., Farenzena, M. and Murino, V.: Multiple-shot person re-identification by hpe signature, *Proc. ICPR*, pp. 1413-1416 (2010).
- [4] Bak, S., Corvee, E., Bremond, F. and Thonnat, M.: Multiple-shot human re-identification by mean riemannian covariance grid, *The 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), 2011*, IEEE, pp. 179-184 (2011).
- [5] Metternich, M. and Worring, M.: Semi-interactive tracing of persons in real-life surveillance data, *Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence*, ACM, pp. 43-48 (2010).
- [6] Fischer, M., Ekenel, H. and Stiefelhagen, R.: Person re-identification in TV series using robust face recognition and user feedback, *Multimedia Tools and Applications*, Vol. 55, No. 1, pp. 83-104 (2011).
- [7] Gray, D., Brennan, S. and Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking, *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS), 2007.*, Vol. 3, p. 5 (2007).
- [8] Rocchio, J.: Relevance feedback in information retrieval (1971).
- [9] オムロン株式会社: OKAO Vision.
- [10] Farenzena, M., Bazzani, L., Perina, A., Murino, V. and Cristani, M.: Person re-identification by symmetry-driven accumulation of local features, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010*, IEEE, pp. 2360-2367 (2010).
- [11] Balzano, L., Nowak, R. and Recht, B.: Online identification and tracking of subspaces from highly incomplete information, *The 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2010.*, IEEE, pp. 704-711 (2010).