

# 時空間を横断した出現パターンに基づく高速な映像検索

劉 健全<sup>†</sup> 西村 祥治<sup>†</sup> 荒木 拓也<sup>†</sup>

<sup>†</sup> 日本電気株式会社グリーンプラットフォーム研究所 〒 211-8666 神奈川県川崎市中原区下沼部 1753

E-mail: †{j-liu@ct, s-nishimura@bk, t-araki@dc}.jp.nec.com

あらまし 本論文では、防犯カメラが撮影した映像から、同一人物がいつ・どこで現れたか（出現パターン）による検索技術を紹介する。本稿で紹介する技術は、カメラ映像から「顔」を抽出し、顔の類似性から同一人物とみなせる出現を抽出することで高速な映像検索を実現する。これにより、複数の事件現場に共通して現れた不審者の特定など、検索対象の顔が不明なため従来困難であった犯罪捜査・予防を実現する。このような検索を行なうためには既存技術では、すべてのシーンの出現に対して「同一人物とみなせるほど顔が類似しているかどうか」を確認する必要がある。これをすべての人に対して行わなければならないため、膨大な数の照合が必要となり、非常に長い時間が必要となる問題が生じる。これに対して、我々は独自のツリー状のデータ構造を用い、お互いに類似したデータを索引することにより、高速な映像検索を実現した。評価実験により提案技術は既存技術より100倍程度の高速化が可能であることを検証した。また、デモにより高速化の様子を示す。

キーワード 類似検索, 映像検索, データ索引, 出現パターン, 類似分類

## 1. はじめに

近年、ショッピングモール、ビル、駅、公共施設など至るところに防犯カメラが設置されるようになってきた。従来より、防犯カメラは、警備や捜査といったパブリックセーフティの用途として利用されている。その際、映像の確認や解析は、主に人手によって行なわれている。しかしながら、昨今、防犯用途の普及によりカメラの設置台数は急激に増加し、カメラから集まる映像データも増大している。2020年には、防犯カメラが撮影した映像が5.6 ZBに達し、ビッグデータ全体の42%を占めると予測される[1]。このため、人手による解析が困難になってきている。例えば、複数箇所に設置されたカメラの映像を対象に、人の目による確認作業では多くの時間が必要であるため、同じ場所に何度も出現する、あるいは複数の場所に出現する人物などの検索をおこなうのは、非常に困難である。

このような人的資源の限界を克服するため、映像に写った人物、車両、物体をリアルタイムに自動的に解析する技術が開発されてきた[2]。例えば、最新の顔認識技術は、他人許容率が0.1%の時、誤照合率が2-4%という高い精度で人物を識別可能である[3]。映像監視に関する研究が過去20年に渡って行なわれてきた[4]が、多くの既存技術は典型的な映像解析、例えば、物体の発見、認識、追跡、および行動分析に留まり、大量の防犯カメラを横断した映像検索技術は今なお欠けている[5]。そこで、我々は大量の防犯カメラ向けの映像検索技術に着目し、指定した顔の人物を高速に検索できる映像検索システム[6],[7]を開発した。

しかしながら、パブリックセーフティの用途にある警備や捜査では、必ずしも検索対象を指定できるとは限らない。すなわち、不審者の対象となりうる候補を自動的に発見し、いかに結果を高速に絞り込むことは依然困難な研究課題として残ってい

る。本稿は、このような課題に着目し、時空間を横断した出現パターンに基づいた高速な映像検索技術を提案する。本提案技術は、映像に写り込んだ人物の顔特徴の類似性から同一人物であることを特定し、その人物が異なる時間と空間で頻繁に現れるパターンに基づき、出現頻度の高い人物を不審者の対象として高速に検索して見出す。本技術の主な着眼点としては、同一人物が異なる時間と空間を横断した頻繁な出現にある。例えば、駅や観光地等でターゲットを探しているスリのような人物は、異なる時間に不自然に頻繁に現れるという特徴を持っている。また、複数の放火事件現場に共通して現れた人物も、放火犯などの不審者として疑わしい可能性がある。なぜならば、普通の人にはそもそも異なる放火事件の現場を訪れることは考えにくいからである。

本論文では、このような観点において、防犯カメラが撮影した映像から、同一人物がいつ・どこで現れたかを出現パターンとして扱う。これにより、複数の事件現場に共通して現れた不審者の特定など、検索対象が不明であるため従来困難であった犯罪捜査・予防を実現する。

この実現に向けて2つの重要な課題を解決しなければならない。1つ目は、別のシーンに現れた人物が同一人物とみなせるかどうかを高精度で判断することにある。2つ目は、十分に顔が似ている人物を同一人物のグループに高速に振り分けることにある。前者に関しては、顔照合の速度と精度が世界一<sup>(注1)</sup>と認められた顔認証技術のNeoFace<sup>®</sup><sup>(注2)</sup>を用いることで、顔が十分に似ているかどうかを判断できる。しかしながら、後者に関しては、既存技術では、図1に示すように顔照合を全てのシーンに映っている全ての人に対して行わなければならない

(注1): [http://www.nist.gov/customcf/get\\_pdf.cfm?pub\\_id=905968](http://www.nist.gov/customcf/get_pdf.cfm?pub_id=905968)

(注2): <http://www.nec.co.jp/soft/neoface/product/neoface.html>

め、膨大な数の照合が必要となり、計算量が非常に高い問題が生じる。

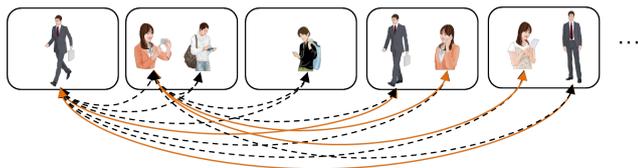


図 1 既存技術での課題: あるシーンに映っている人が他のシーンに映っているか調べるには、大量の照合が必要

このような計算量が問題でなければ、 $N$  対  $N$  照合を要するシーケンシャルスキャン、クラスタリング技術である DB-SCAN [8], [9] と階層的クラスタリング [10] を用いることで、同一人物のグループ振り分けを実現できる。しかし、これらの既存技術は、いずれも  $O(N^2)$  以上の計算量を必要とし、クラスタリング技術がオフラインのアプローチであるため、リアルタイムな映像監視にとっては実用性が低い。よって、本研究では、先行研究 [11], [12] で提案した独自のツリー状のデータ構造を用い、お互いに類似したデータをリアルタイムで動的に索引付けすることにより、高速な映像検索を実現する。

本論文は以下の通りで構成される。2 章で我々の着眼点と問題設定を述べる。3 章で関連研究を紹介する。そして、4 章で我々独自の手法を提案し、5 章において評価実験を用いて提案手法の有効性と効率性を示す。最後に、開発した映像検索のデモシステムを紹介し、全体と今後の課題についてまとめる。

## 2. 問題設定

本章は、本論文が対象とした映像検索に関する問題設定を述べる。まず、検索に用いられる映像の属性情報と人物の特徴情報は以下である。

[映像の時空間情報] 時間情報は、防犯カメラが撮影した時刻である。この時刻は、属性として映像の各フレームに付与される。空間情報は、防犯カメラが設置した場所の位置情報を用いる。多くの場合は位置情報はカメラ ID と紐付いているので、本稿では、簡単にカメラ ID を用いて映像のユニーク性を識別する。

[人物の特徴情報] 本稿では、映像から抽出される顔の特徴量を人物の特徴情報として用いる。顔の抽出および照合は、顔認証技術 NeoFace<sup>®</sup> により行う。顔の照合では、2 つ顔の類似度合いを表すスコア ( $[0, 1]$  区間にある実数) を算出する。

次に、本稿で対象とした映像検索と、前述の時空間を横断した出現パターンについて説明する。

[映像検索] 従来の映像検索は、目的人物の画像をとして与えられた人物の顔などの特徴を用いて、該当人物が映った目的の映像を探し出し、検索対象が明確な検索である。一方、本研究では、防犯カメラが撮影した監視映像による犯罪予防・迅速な捜査の実現を目的としているため、検索対象とする人物はクエリの発行時には不明である。

[出現パターン] 本稿では、複数の防犯カメラ (空間軸) に現れ、

異なる時間 (時間軸) に頻繁に現れる特徴を時空間を横断した出現パターンと呼ぶ。例えば、犯罪予防の用途において、駅や観光地等でターゲットを探しているスリとして疑われる人物は、不自然に頻繁に現れるという特徴を持っている。また、迅速な犯罪捜査の用途において、複数の放火事件現場に共通して現れた人物は、放火犯等の不審者として疑わしい可能性がある。これらの不審人物が持つ共通点は、複数の防犯カメラ (空間軸) に現れ、異なる時間 (時間軸) に頻繁に現れる特徴である。

そして、前述の出現パターンに基づいて、同一人物とみなせる出現を抽出し、同一人物の出現頻度でランキングすることで、本論文が対象とした映像検索を実現した。その高速な実現については、4. で改めて紹介する。

[検索の入力] 本稿の映像検索は、検索対象が不明であるため、検索の入力はないが、検索を実現するには、前述で仮定した出現パターンにより不明人物をグループに振り分け、ランキングを行う。

[検索の出力] 検索後に、時空間を横断した出現パターンを持つ人物の出現頻度順でランキングした人物の上位  $K$  件の一覧を結果として出力する。

## 3. 関連研究

本章は、これまで説明した映像検索問題を解くことが可能な既存技術を紹介する。解決可能な手段は、概ね 2 通りある。 $N$  対  $N$  の全件照合において同一人物に同じ番号 (ID) を振って、異なる ID の人物グループに分けるというシーケンシャルスキャン法 (以降 SeqScan 法) がその 1 つである。もう 1 つは、既存のクラスタリング技術を適応し、蓄積した映像から抽出した人物を異なるクラスタに振り分けるというクラスタリング法 (以降 Cluster 法) になる。これらにより、同一人物のグループに振り分けられた後、同一人物の頻度順で人物をランキングする。上位  $K$  件を結果として出力する。以下に、この 2 通りの既存技術を述べる。

SeqScan 法では、リアルタイムに映像から人物の顔を抽出し、顔の特徴量を抽出する。抽出された顔の特徴量に顔 ID を付与する。最初の 1 つめは、ID を 1 とする。次に入ってくる顔の特徴量を既に ID を付与した顔の特徴量と照合する。指定したスコアのしきい値を超えた場合は、該当顔と十分に類似すると判断し、同じ ID を付与する。この照合を繰り返しおこなうことにより、全ての顔が同じ ID を持つ同一人物のグループに分類される。そして、グループごとにおいて、同一人物を時間軸と空間軸で出現パターンとしてまとめる。最後に、各グループの出現頻度順でランキングする。

SeqScan 法は、同一人物を判定しながら、番号を振ることが可能であるため、リアルタイムまたはオフラインに対応できる。しかしながら、ID を振るためにかかる特徴量間の照合の数が膨大になり、計算量  $O(N^2)$  がかかる。長時間の映像検索にはスケールしない問題が生じる。

Cluster 法では、特徴量の抽出およびグループのランキングを SeqScan 法と同様に処理するが、全ての顔を同一人物のグループに振り分ける部分のみは、既存のクラスタリング技術を

適応することで実現する．例えば，密度に基づくクラスタリング手法の DBSCAN [8], [9] と，階層的なクラスタリング手法の Agglomerative アルゴリズム [10] が挙げられる．しかしながら，DBSCAN 法は  $O(N^2)$ ，Agglomerative 法は  $O(N^3)$  の計算量がかかり，いずれも実用的ではない．しかも，クラスタリング処理では，対象となる全データをためてからクラスタリングを行わなければならないため，オフライン解析のみ対応できる．

本研究が応用の目的とした迅速な犯罪捜査とリアルタイムな分析による犯罪予防にとって，Cluster 法は実用ではないため，実験では SeqScan 法との比較により評価を実施した．

#### 4. 提案手法

本章は，先行研究 [11], [12] で提案した独自のツリー状のデータ構造を用い，お互いに類似したデータをリアルタイムで動的に索引することにより，独自の高速な映像検索手法について紹介する．

先行研究 [11], [12] では，従来の映像検索を高速化するために，図 2 に示した独自のツリー状のデータ構造を提案した．この独自の木構造を類似度木と呼ぶ．類似度木の構築または性能評価については，先行研究 [12] で紹介したため，本稿では，その特徴と独自性について簡単に説明する．

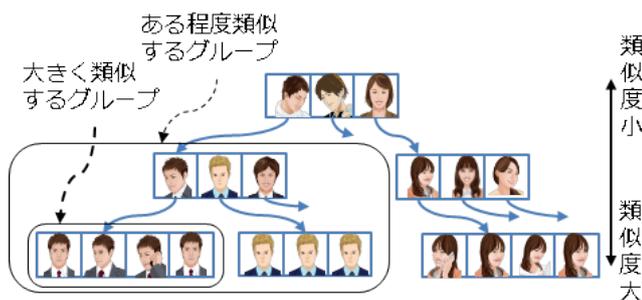


図 2 独自のツリー状のデータ構造 (類似度木)

類似度木では，データを類似度に基づき木構造で管理する．図 2 に示したように，下の層になるほどお互いに類似したデータが集まり，上の層になるほどお互いに異なったデータになるようにデータを配置する．この類似度木を用いることにより，お互いに類似したデータを抽出する際には，抽出したい類似の程度によって，下の層からデータを取り出すことで，容易にデータのグループを再構築できる．

また，類似度木はリアルタイムで構築することができる．データを挿入するたびに木構造を更新し，前記の性質を常に保つように動作する．木構造の更新にかかる時間は極めて小さいため，大量の防犯カメラ映像から得られる大規模な顔情報をリアルタイムに挿入することが可能になる．これにより，防犯カメラが撮影した直後のデータについても即時に検索可能になる．

次に，類似度木を用いたデータのリアルタイムな類似分類を実現するアルゴリズムを提案し，その詳細について説明する．

アルゴリズム 1 では，抽出したい類似度合いのしきい値  $\delta$  を指定し，リアルタイムに構築された類似度木  $T$  を用いて，デ

---

#### Algorithm 1: GroupBy( Threshold $\delta$ , SimTree $T$ )

---

**Input:**  $\delta$ , specifying the minimum similarity of two data.

**Input:** SimTree  $T$ , built by BuildIndex [12] algorithm.

**Output:**  $G$ , a list of groups constructed by this algorithm.

---

```

1 begin
2   Queue  $Q \leftarrow \emptyset$ ;
3   foreach Node  $e \in T$  do
4     if  $e.\delta \geq \delta$  then
5        $Q.inqueue(e)$ ;
6   SimTree  $T2 \leftarrow \emptyset$ ;
7   while  $Q$  is not empty do
8     Node  $e \leftarrow Q.dequeue()$ ;
9     foreach Data  $d \in e$  do
10      if  $SimilaritySearch(d, \delta, T2) \neq \emptyset$  [12] then
11         $G.insertToExistingGroup(d)$ ;
12      else
13         $G.insertToNewGroup(d)$ ;
14       $T2.insertData(d)$ ;
15  return  $G$ ;

```

---

ータ間の類似度が  $\delta$  以上であるものを同一グループに振り分ける処理を示す．行 2 から 5 において，既に構築された類似度木  $T$  を辿り，類似度が  $\delta$  以上のノードを切り出してキュー  $Q$  に入れておく．そして，行 6 から 14 において，キューに入ったノードを取り出し，ノードに保存される全てのデータを 1 個ずつ新しい類似度木  $T2$  に挿入する．データの挿入するたびに，類似検索をかけて，該当データと類似したデータが存在するかどうかを検査する．存在する場合は，既存グループに振り分け，存在しない場合は，新規グループに振り分ける．最後 (行 15) に，振り分けられたグループをアルゴリズムの結果として返す．

最終的に，本稿で限定した映像検索をおこなうたびに，アルゴリズム 1 を呼び出し，振り分けられたグループごとに時空間を横断した出現パタンの頻度順で人物をランキングする．

また，アルゴリズム 1 により抽出されるグループは，従来の結果と比べ近似結果となるが，理想の場合は計算量が  $O(M \log(M))$  に近づくため，本稿で限定した映像検索を高速に実現できる．ここでは， $M$  がデータベースに入った人物のユニーク数を仮定する．

#### 5. 実験と評価

本章では，実データおよび模擬データを用いて，本稿で提案した映像検索の性能を評価した．評価実験は，メモリ 11GB と Intel<sup>®</sup> Xeon<sup>®</sup> E5504 @ 2.00GHz の CPU を持つマシン環境において，Ubuntu 12.10 の OS 上で実施した．表 1 に記載した実験データにより，しきい値  $\delta = 0.8$  をデフォルト値として映像検索の速度と，結果の再現率と精度を評価した．

##### 5.1 データセット

本実験では，模擬データと実データの 2 種類データを用いて実証実験を実施した．

表 1 実験データセット

名前	サイズ
模擬データ	のべ 15 分の模擬撮影映像から得た 1 万件の顔特徴量
実データ	のべ 24 時間のカメラ映像から得た 100 万件の顔特徴量

模擬データは、エキストラを用いて、複数カメラで重要施設の周辺に撮影した模擬映像から得た 1 万件の顔特徴量である。模擬データにおいて、既知の不審者役を演じた 1 人が、特定の場所をうろろする行動を意図的におこなった。

実データは、海外の特定地域の公的機関の協力を得て、実際の街角の防犯カメラ映像のべ 24 時間を用いて、100 万件の顔特徴量を抽出した<sup>(注3)</sup>。実データにおいて、我々に知らされていない 7 人の「不審者役」が、公的機関によって動員され、彼らは複数箇所の対象エリアを頻りに訪れるなど、不審な行動を意図的におこなった。また、不審者役の顔、人数を伏せた状態で評価をおこなった。

### 5.2 評価結果

模擬データを用いて、本稿で提案した映像検索技術で解析したところ、不審者役を演じた 1 人は、上位 10 人の検索結果の中に含まれ、1 位にランクされた。検索結果の再現率は 100% に達し、Top@1 の検索精度も 100% に達したことが分かった。図 3 は、模擬データにおいてデモシステムが実行した映像検索の結果を示した画面である。

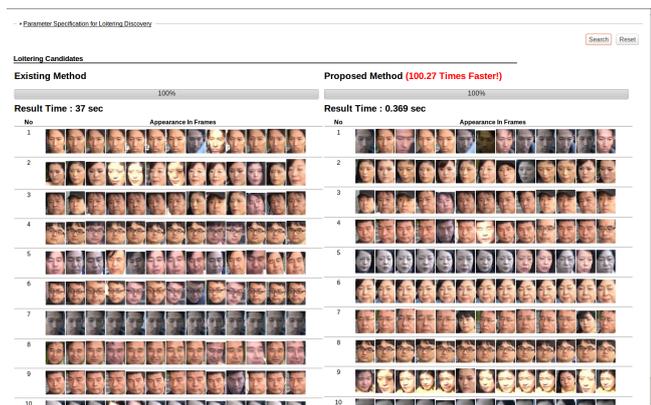


図 3 模擬データにおいて映像検索を実行したデモシステムの画面。既存技術 SeqScan 法の実行速度より 100 倍向上。

図 3 は、本稿で提案した技術と既存技術の SeqScan 法を実装したデモシステムを紹介する。左側は、既存技術の SeqScan 法を用いて実行した検索結果を示す。右側は、本提案技術を用いて実行した検索結果を示す。既存技術と同様に、提案技術は不審者役を演じた 1 人を検索結果の 1 位にランクした。既存技術を用いた検索でかかった 37 秒と比べ、提案技術はわずかの 0.369 秒で完了し、実効速度を 100 倍向上した。

また、実データを用いても、同様な結果を得ることができた。本稿で提案した映像検索技術で解析したところ、同じ場所に長時間滞在したり、頻りに現れたりする“不審”な人物の検索を、わずか 10 秒という短時間で完了させた。検索結果で「怪しい」

と思われる上位 20 人を抽出し、公的機関に確認してもらったところ、7 人の「不審者役」は全員、この 20 人の中に含まれていた。

実データを用いた実験においても、本提案技術は、24 時間の映像を検索する時間を従来の SeqScan 法でかかる約 1 時間から 10 秒まで大幅に短縮した。公的機関を介した検証結果により、検索結果の再現率は、100% に達したと言える。不審者役 7 人の顔が知らされてないため、検索結果の精度は評価しなかった。

## 6. まとめ

本稿では、防犯カメラが撮影した映像から、同一人物がいつどこで現れたかという出現パターンに基づき、独自のツリー状のデータ構造を用いた類似データの索引により、検索対象が不明な映像検索を高速に実現した。評価実験により提案技術が既存技術より 100 倍程度の高速化であることを検証した。

今後の課題として、映像以外の実データを用いてさらに実証実験を実施することと、近似検索アルゴリズムを改善して正確かつ高速な映像検索手法を検討することにあると考えられる。

## 文 献

- [1] John Gantz and David Reinsel. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. *IDC Technical Report*, Dec. 2012.
- [2] Mubarak Shah, Omar Javed, and Khurram Shafique. Automated visual surveillance in realistic scenarios. *IEEE Transactions on Multimedia*, 14(1):30–39, 2007.
- [3] 越仲孝文, 大網亮磨, 細見格, 今岡仁. 音声・映像認識連携への取り組み: 1. 音声・映像情報の構造化と検索. *情報処理*, 52(1):71–78, Jan. 2011.
- [4] Tomi Rätty. Survey on contemporary remote surveillance systems for public safety. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(5):493–515, 2010.
- [5] Weiming Hu, Dan Xie, Zhouyu Fu, Wenrong Zeng, and Stephen J. Maybank. Semantic-based surveillance video retrieval. *IEEE Transactions on Image Processing*, 16(4):1168–1181, 2007.
- [6] 西村祥治, 劉健全, 藤森偉恭, 荒木 拓也. Wally: 映像検索システムを対象としたスケーラブルな分散データストア. In 第 5 回データ工学と情報マネジメントに関するフォーラム (DEIM2013), pages P4–1, 2013 年 3 月 3 日–3 月 5 日.
- [7] Jianquan Liu, Shoji Nishimura, and Takuya Araki. Wally: A scalable distributed automated video surveillance system with rich search functionalities. In *Proceedings of the 22nd ACM MM*, pages 729–730, 2014.
- [8] Martin Ester, Hans peter Kriegel, Jrg S, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd ACM SIGKDD*, pages 226–231, 1996.
- [9] Junhao Gan and Yufei Tao. DbSCAN revisited: Mis-claim, un-fixability, and approximation. In *Proceedings of the 36th ACM SIGMOD*, pages 519–530, 2015.
- [10] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques*, chapter 7: Cluster Analysis. 3rd edition, 2011.
- [11] 劉健全, 西村祥治, 荒木 拓也. 類似度の階層関係に基づく木構造索引を用いた効率的な類似検索. In 第 5 回データ工学と情報マネジメントに関するフォーラム (DEIM2013), pages A9–1, 2013 年 3 月 3 日–3 月 5 日.
- [12] 劉健全, 西村祥治, 荒木拓也, 中村 祐一. 類似度の階層関係を用いた類似検索の高速化とその応用. *電子情報通信学会技術研究報告*, 115(315):123–128, November 2015.

(注3): [http://jpn.nec.com/press/201511/20151111\\_02.html](http://jpn.nec.com/press/201511/20151111_02.html)