映像データベースのコアである類似索引技術とその新しい応用

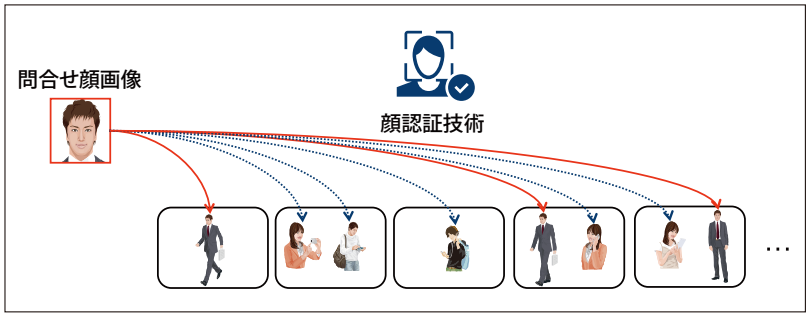
西村 祥治　劉 健全　NEC

なぜ今，映像データベース技術に注目すべきなのか

防犯カメラの普及により，膨大な映像が記録，蓄積されるようになってきている．しかしながら，蓄積される映像の量が膨大であるがゆえ，そのすべてに人が目を通すことは事実上不可能である．このため，関心がある内容が映っている瞬間の画像（シーン）を即座に探し出す映像検索システムに注目が集まっている．このような映像検索システムを支えるのが画像認識技術と映像データベース技術である．画像認識技術はいわば機械の眼に相当する．近年深層学習技術の発達により，認識可能な対象が拡大するとともに，認識精度が実用レベルに近づいている．一方，映像データベース技術は映像を整理整頓する技術である．これにより，大量の映像の中から所望のシーンを瞬時に取り出すことが可能になる．従来の映像データベース技術では検索の高速性や大規模化がその役割の中心であった．しかし，今後は新しい映像の探し方など映像検索の高度化を切り開く役割が大きくなると考える．本稿では，映像検索システムを概観し，映像を整理整頓するためのコア技術である類似索引技術について解説をする．そして，この類似索引技術を活用することで実現した最新の映像検索技術である時空間データ横断プロファイリング[1]を紹介する．

映像検索システムの概要

映像検索とは，大量の映像から目的の映像を探し出すことである．本稿では，内容に基づく検索（Content-based Retrieval）と呼ばれる検索技術を対象とする．図-1に顔認証技術を用いた映像検索の例を示す．これはテレビドラマや映画などで指名手配中の人物を顔写真から探す場面を想像してもらえるとよい．入力として顔画像を与え，顔認証技術により映像に映った顔画像を照合し，合致した顔画像が映ったシーンを探し出す．

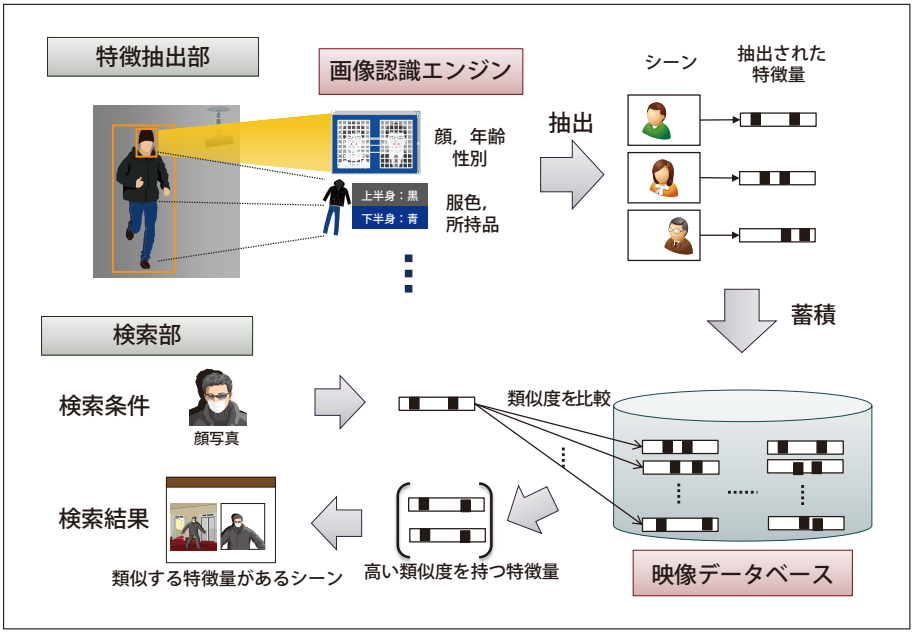


■図-1　内容に基づく検索（顔画像を用いた検索）

映像検索システムはおおむね図 -2 のように構成されるのが一般的である．映像検索システムは大きく特徴抽出部と検索部の 2 つに分けられる．特徴抽出部は映像を構成する画像一枚一枚に対して，画像認識エンジンを用いて，その画像に含まれる特徴量と呼ばれるデータを抽出する．そして，特徴量とその特徴量が出現した画像（シーン）と関連付け，映像データベースに格納する．検索部は，問合せ（クエリ）を入力として受け取り，その問合せに対応する特徴量に変換する．そして，その特徴量をキーとしてデータベースで照合し，その特徴量に関連付けられたシーンを結果として出力する．

映像データベースで特徴量を管理する方法は，画像認識エンジンが出力する特徴量の種類に大きく依存する．画像認識技術は大きく一般物体認識と特定物体認識の 2 種類がある．一般物体認識とは物体が属するクラスに振り分けて認識する技術である．たとえば，映像に映った物体を人，犬，車，カバンなどとして認識する．特定物体認識は，個体（インスタンス）を判別する技術である．典型的な例は顔認証技術で，個人ごとに顔を認識する．データの型という観点で見ると，一般物体認識の結果はクラスを表すタグのようなシンボルデータとして表現されることが多い．一方，特定物体認識の結果は特徴ベクトルと呼ばれるビット列データとして表現されることが多い．また，同一の対象同士であっても，画角，色合いなどの環境変動があるため，ビット列データ同士が完全に一致することはほとんどない．このため，データ同士の類似の度合いを表す類似度と呼ばれる数値を返す照合関数が提供される．そして，類似度が一定の閾値を超えたかどうかによって，一致・不一致を判定する．

特徴量がシンボルデータである場合は，シンボル同士が完全一致するかどうかを効率的に検査できればよいので，一般的なリレーショナルデータベースを用いた管理で十分である．一方，特徴量がビット列データであり，照合関数で類似度を計算しなければならない場合，管理方法に工夫が必要となる．基本的な操作である類似検索を例にどのような工夫が必要であるかについて説明する．類似検索とは，図 -1 にあるように問合せ顔画像と照合し，その類似度が閾値を超えたものを探すことである．全件について照合すれば，所望のシーンをすべて得ることが可能である．しかし，件数が増加するほど，その照合にかかる時間も線形で増加する．このため，検索時間を短くするためには照合回数をできる限り少なくしなければならない．その一方で，照合されないデータが増えることで，本来は結果として列挙すべきデータが結果から漏れることが発生する．この漏れ具合を評価する指標を再現率といい，漏れがまったくなかっ



■図 -2　映像検索システムの一般的な構成

た場合を 100%，すべて漏れていた場合を 0% とする．検索時間と再現率との間にトレードオフの関係があり，双方を高い水準で保てるようにデータを管理することが映像データベースにおける中心的な研究課題となっている．

映像検索を支える類似索引技術

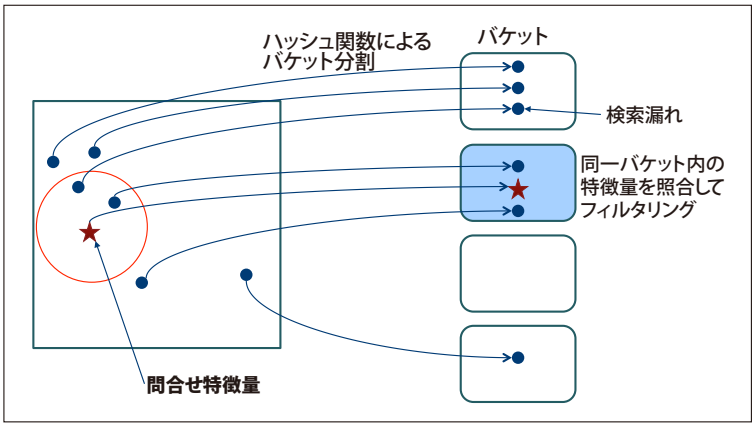
検索時間をできるだけ短く，再現率をできるだけ高くするためのデータ管理技術が類似索引技術である．類似索引技術は問合せデータに合致しそうなデータだけを選択的に照合するための仕組みを実現する．類似索引技術はデータ構造から大きくハッシュ型とツリー型に分けることができる．以下では代表的な類似索引技術について紹介する．

ハッシュ型類似索引

ハッシュ型類似索引は，類似するデータ同士が高い確率で同じハッシュ値を持つような特殊なハッシュ関数を用いたデータ構造である．代表的なものとして局所性鋭敏型ハッシュ（LSH, Locality Sensitive Hashing）[2] がある．

LSH の基本動作について図 -3 を用いて説明する．まず，特徴抽出部により抽出された特徴量に対してハッシュ関数を用いて，ハッシュ値ごとにバケットに分割して管理する．次に，問合せがあったとき，問合せの特徴量にハッシュ関数を用いて，そのハッシュ値に対応するバケットにアクセスする．そして，そのバケットにあるすべての特徴量と照合関数を適用し，類似度が閾値を超える特徴量を結果として返す．このとき，照合回数は問合せ特徴量と同じハッシュ値を持つ特徴量の数に抑えることができる．一方で，問合せ特徴量と照合した結果，閾値を超えるにもかかわらず，異なるハッシュ値を持つ特徴量は，結果から漏れることになる．再現率を改善する方法として，複数の異なるハッシュ関数を用いる方法が知られている．すなわち，ハッシュ関数ごとに特徴量をバケット分割しておき，検索時にハッシュ関数ごとにハッシュ値が一致したバケットにある特徴量に対して照合する．ハッシュ関数の数を増やすにつれ，照合対象の特徴量が増えるため再現率は改善する．一方で，照合回数が増えるので検索時間は悪化，また，バケットの数も増えるため空間使用量も悪化する．

実用上の課題は，特徴量に適合したハッシュ関数を構成することが難しいことである．LSH を提案した Indyk と Motwani は，上記のような性質を持つハッシュ関数の構成方法として Hamming distance を用いた方法と安定分布を用いた方法を提案している．すなわち，前者は，特徴量をビット列とみてランダムに一部のビットだけを抜き出すことでハッシュ関数を構成する．後者は，特徴量を n 次元のベクトルと見て，安定分布に従った乱数を要素に持つベクトルとの内積をとることで数値に写像する．これらの構成方法は，特徴量のデータ形式や照合関数の性質を仮定して設計されている．このため，Web などで公開されている画像認識エンジンを入手して利用する場合，特徴量に適用するハッシュ関数が想定している仮定を満たしているかどうかに注意を払う必要がある．



■図 -3　ハッシュ型類似索引の一例（LSH）

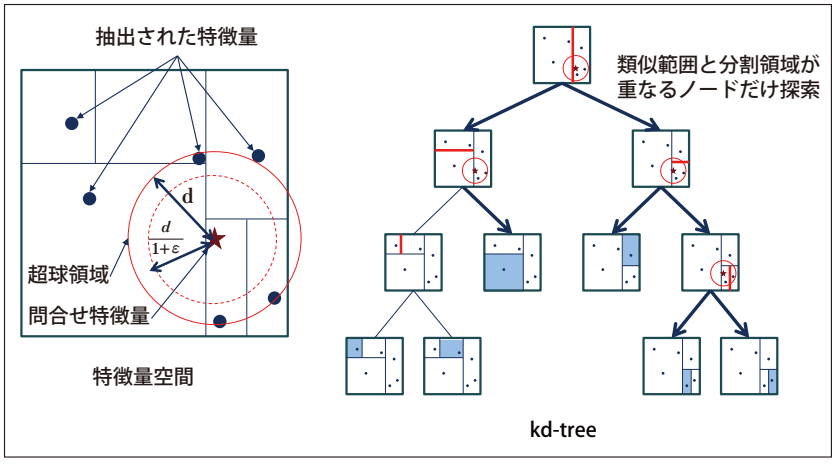
ツリー型類似索引

索引を構成するもう一つの方法は木構造（ツリー）を用いることである．基本的なアイディアはデータを階層的に分割し，類似するデータ同士がなるべく木構造上で近傍になるように配置する．データをどのように再帰的に分割するかによって，空間をベースに分割する方法と，データ同士の類似性に基づいて分割する方法に分類することができる．

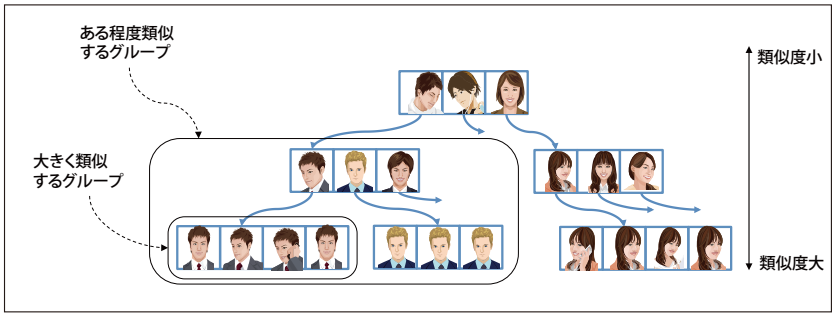
空間をベースに分割する方法として代表的なものとして，kd-tree をベースにした ANN（Approximate Nearest Neighbors）[2] がある．この方法では特徴量は k 個の数値からなる多次元ベクトルデータであることを仮定する．kd-tree は多次元データを扱う木構造で，階層ごとに空間の軸を 1 つ選び，空間を 2 分割する．最も単純な 2 次元の場合を例として挙げて説明する．図 -4 にあるように 6 つの特徴量があるとする．そして，空間分割は，階層ごとに垂直方向，水平方向の順で再帰的に繰り返しながらデータを二分する位置に軸をとることで行う．類似検索は，問合せ特徴量を中心とした類似半径 d で囲われる超球領域と kd-tree の各ノードに対応する部分空間との重なりの有無により枝刈りをすることで実現できる．しかし，高次元になるほど球面集中効果により超球領域にほとんどの部分空間が重なりやすくなるため，枝刈りがほとんど効かなくなる．そこで，ANN は d よりも $1/(1+\varepsilon)$ 倍の半径を用いて球面集中効果を緩和し，高速化を図っている．一方で，d よりも小さい半径を用いていることから検索漏れが発生するため，再現率は悪化する．

データ同士の類似性に基づいて分割する手法として Luigi [3] がある．Luigi は，階層ごとに類似の閾値を設定し，階層が深くなるほどその閾値を厳しくするように構成されたデータ構造である（図 -5）．特徴量を木構造のルートノードからデータを挿入し，そのノードにすでにある特徴量と比較し，その類似度が閾値よりも類似しているならばその特徴量の子ノードへ，そうでないならそのノードに配置するという操作を再帰的に繰り返す．類似の閾値を階層が深くなるほど厳しくするため，類似する特徴量同士が自然と同じノードに集まりやすくなる．Luigi では特徴量のデータ分布に従って階層が深くなることがあるため，階層の深さはバランスしない．このため，特徴量のデータ分布が極度に偏っている場合，線形検索に陥る可能性がある．

Luigi の特徴の一つは，照合結果である類似度の情報しか用いないため，特徴量や照合関数の詳細が



■図 -4　kd-tree を用いた ANN（Approximate Nearest Neighbors）



■図 -5　ツリー型類似索引の一例（Luigi）

ブラックボックスであってもよい点である．これは，汎用性や実用上，重要な性質になる．

以上，映像データベース技術のコアである類似索引技術について紹介した．これらはほんの一部であり，これらの手法を組み合わせたハイブリッドな手法も多く提案されている．また，検索の速度，精度のトレードオフだけでなく，空間使用量，汎用性など工夫や改良の余地はまだまだ広い．このため，今後もさまざまな手法が提案されることが期待される．この分野に興味を持った方には，画像検索向けの類似索引技術に関する記事がまとまっている参考文献2）をおすすめする．また，より網羅的に知りたい場合は多次元索引技術の大著である文献4）を参照するとよい．

事例紹介：時空間データ横断プロファイリング

最後に映像検索の最新事例として，不特定の人物を対象にした映像検索技術である時空間データ横断プロファイリング[1]を紹介する．

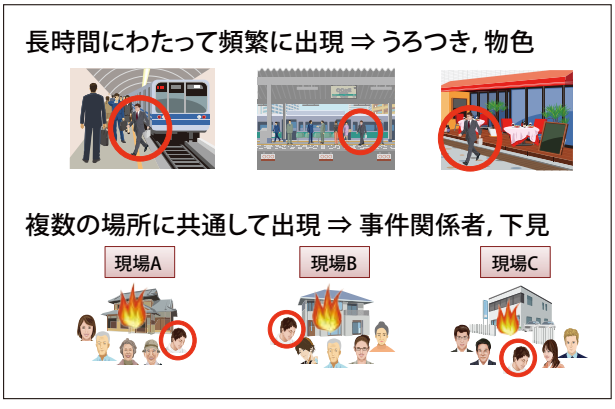
従来の映像検索は，図-1のように探したい対象を入力として与え，それが出現するシーンを結果として返す．このため，事前に入力となる画像を入手できない場合，検索することができない．

そこで，顔画像を指定せず，映像中，何度も出現するなど特徴的な出現パターンを示す人物を高速に見つける時空間データ横断プロファイリングを開発した．これにより，図-6のように，スリやうろつきなど複数のカメラに何度も映り込んだ人物や複数の現場に共通して現れる人物を探し出すことができる．

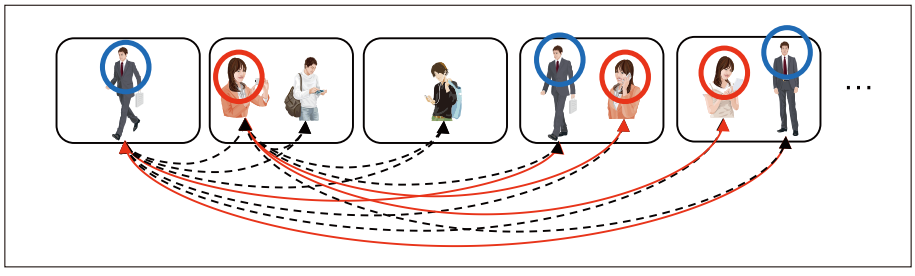
このような検索を実現する際，単純な方法では映像に映ったすべての人物について総当たりで顔照合を実行し，人物ごとに出現回数を集計する必要がある（図-7）．このため，本稿の冒頭で紹介した類似検索と比べても照合回数が桁違いに多い．そこで，前述のLuigiを適用し，類似の階層性を利用することで，この計算コストを劇的に削減することが可能になった．

時空間データ横断プロファイリングによる映像検索例として，複数のカメラに頻出するうろつきの発見を図-8に示す．図-8の左側が検索結果の画面であり，映像中に出現した人物を出現頻度順で列挙している．実はランク1位の人物は不審者役として複数のカメラに映り込んだ人物であり，図-8の右側にその人物が出現したシーンを示す．

これまで，国内外のお客様の協力のもと実証実験



■図-6　時空間データ横断プロファイリングのユースケース



■図-7　単純な方法による時空間データ横断プロファイリングの実現

を行い，雑踏中何度も現れるうろつきの発見，複数の会場での下見に現れた人物の発見，ATM からの不正な出金を行う人物の発見など，不特定の人物を対象とした映像検索の有効性を確認した．たとえば，海外の公的機関の協力を得た実証実験では，複数のカメラから得たのべ 100 万件の顔特徴量に対して，うろつき発見検索を約 10 秒で完了することができた．このとき，技術評価のため協力先は我々に伏せて数名うろつかせていたが，7 名全員を検出することに成功した．

このように，映像データベース技術は従来の映像検索が対象としていた類似検索を高速化するだけにとどまらない．この事例が示すように，アイディアや工夫次第でこれまでにない種類の映像検索を切り開く可能性を秘めている．

展望

今後の映像検索は，画像認識技術と映像データベース技術が両輪となり，より進化すると考える．画像認識技術は機械の眼として，深層学習技術の急速な発展により，多種多様な対象が精度高く見ることができるようになる．そして，画像認識技術により取り出すことができる特徴の種類と量が増加すればするほど，どの特徴に着目し，整理し，要約するかといった使い方が応用上重要になる．今後，映像データベース技術は特徴量を単に管理する技術としてだけでなく，映像を編集，要約する技術としての役割が増えていくのではないかと考える．本稿を通じて，映像データベース技術に興味を持つようになれば幸いである．

参考文献

1) Liu, J., Nishimura, S. and Araki, T.: Visloiter: A System to Visualize Loiterers Discovered from Surveillance Videos, In Proceedings of ACM SIGGRAPH Posters 2016, pp.47:1-2.
2) 八木康史，斎藤英雄編：CVIM チュートリアルシリーズ　コンピュータビジョン最先端ガイド 3，アドコム・メディア（株）(2010).
3) Liu, J., Nishimura, S., Araki, T. and Nakamura, Y.: A Loitering Discovery System Using Efficient Similarity Search Based on Similarity Hierarchy, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 100-A(2):367-375 (2017).
4) Samet, H.: Foundations of Multidimensional and Metric Data Structures, Morgan Kaufmann (2006).

(2017 年 11 月 2 日受付)

西村祥治（正会員）　s-nishimura@bk.jp.nec.com
2001 年京都大学大学院情報学研究科修了．同年 NEC 入社，現在，NEC システムプラットフォーム研究所主任研究員．HPC，並列分散システム，大規模データベース等の研究に従事．

劉　健全　j-liu@ct.jp.nec.com
2012 年筑波大学システム情報工学研究科博士課程修了．同年，NEC 入社．現在，システムプラットフォーム研究所主任．2015 年より法政大学大学院理工学研究科兼任講師．ICSC2018，ISM2017，ICSC2016，BigMM2016 各プログラム共同委員長．博士（工学）．

出現回数が多い人物順にランキング　　ランク1位の人物は不審者役

■図 -8　雑踏中に何度も現れるうろつきの発見