

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

西村 祥治[†] 劉 健全[†] 藤森 偉恭[†] 荒木 拓也[†]

[†] 日本電気株式会社 〒 211-8666 神奈川県川崎市中原区下沼部 1753

E-mail: [†]s-nishimura@bk.jp.nec.com, ^{††}j-liu@ct.jp.nec.com, ^{†††}y-fujimori@bc.jp.nec.com,
^{††††}t-araki@dc.jp.nec.com

あらまし ショッピングモール、ビル、駅など至るところに監視カメラが設置されるようになり、犯罪捜査、迷子探し、遺失物発見などカメラ映像に対する検索システムの重要性が増している。我々は映像から抽出された人物の顔特徴量、服特徴量、年齢、性別といった映像メタデータを元に目的の映像を検索するシステムを開発している。本システムは、大量のカメラから長時間に渡り生成される大規模な映像メタデータを管理し、さらに、ユーザとのスムーズな対話を実現する必要がある。そこで、本論文では、ワークロードの特徴を考慮し、リレーショナルデータベースとメモリキャッシュとを組み合わせたスケーラブルな分散データストアを提案する。実映像を用いた顔検索と服検索によって、本データストアの検索性能およびスケーラビリティを評価する。

キーワード 映像検索, メタデータ, 分散ストア

1. はじめに

近年、ショッピングモール、ビル、駅など至るところに監視カメラが設置されるようになってきた。従来より、監視カメラは、警備や捜査といったパブリックセーフティの要として利用されてきた。その際、その映像の確認や解析は、主に人の目によって行われてきた。しかしながら、監視カメラの設置台数が増加するにつれ、目視による監視は厳しくなりつつある。

一方で、近年の映像解析技術の発達が目覚ましい。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 [1]。そこで、我々は、これらの技術を応用し、映像から抽出された人物の顔の特徴、服の特徴、年齢、性別といった映像メタデータを元に、目的の映像を検索するシステムを開発した。

しかしながら、このような用途の映像検索システムを実用化するにあたり、2つの問題を解決する必要がある。

一つめは、大量のカメラから長時間にわたり生成される大規模な映像メタデータを管理することである。例えば、カメラ 1 台が毎フレーム 1KB の映像メタデータを生成すると仮定すると、1 日で 1.3GB、1 か月で 40GB が生成される。そして、カメラ 1000 台だと 40TB になる。映像そのものと比べると映像メタデータは小さいとはいえ、大勢の人物が映る場合、精度を上げるために特徴量のサイズを増やす場合、様々な解析エンジンを組み合わせる場合などを考えると簡単

にそのデータ量は何十倍にもなる。そして、時間の経過とともにデータ量は増大し続けるため、映像検索システムは、それに合わせてシステム自体がスケールできる必要がある。

二つめは、ユーザとのスムーズな対話を実現する必要がある。例えば、不審人物を見つけたとき、そのカメラだけでなく、その周辺のカメラにもその人物が映っていないかを即座に調査したい。この時、検索に何時間もかかっているのは取り逃がす可能性がある。短時間で応答するには、検索性能が十分に高速であることはもちろん、映像が撮影されると同時に検索可能な状態にする必要もある。

これらの問題を解決するために、我々は監視システムにおける映像検索ワークロードの特徴を考慮した、スケーラブル、かつ、高速な検索を可能にする分散データストア Wally を開発した。Wally は、リレーショナルデータベースとメモリキャッシュとを組み合わせたスケールアウト型のデータストアである。そして、我々は、監視映像において典型的なワークロードである最近撮影された映像の検索する場合と長時間にわたり撮影された映像をくまなく検索する場合の両方に対応できるデータ配置アルゴリズムも開発した。これにより、前者の場合は応答時間指向で、後者の場合はスループット指向で検索することが可能である。

本論文の構成は、以下のとおりである。2 章では、映像検索システムを構成する上での関連研究について述べる。3 章では、映像検索システム向けのメタデータストア Wally の設計と実装について述べる。4 章では、映像検索システムによ

るメタデータ検索の具体例を紹介する。5章では、メタデータストア Wally を性能評価する。6章では、全体と今後の課題についてまとめる。

2. 関連研究

大規模なカメラで撮影された大量の映像に対する映像検索システムは、複数の要素技術を組み合わせることで構成される。

何よりもまず、映像からメタデータを抽出する画像解析技術である。映像から人物の顔や服の特徴を抽出する解析技術は、すでに実用化され、製品化されている [1], [2]。さらに、人物の年齢や性別まで推定する製品 [3] も存在している。

次に、大規模なカメラからメタデータを抽出するには、これらを効率的に解析処理する技術が必要である。例えば、文献 [4], [5] は、写っているシーンの状況に応じて解析処理を適切にスケジューリングする技術を提案している。

最後に、効率的な検索性能を実現するためには、メタデータ管理技術が必要である。大規模なカメラシステムを対象とし、長期に渡りメタデータを蓄積する場合、メタデータ量は膨大になる。画像解析と画像照合は組となる技術であることから、従来より照合の高速化手法も提案されてきた。例えば、文献 [6] は、近似最近傍検索を用いて、大量の顔照合の高速化手法を提案している。

一方で、本論文が対象とするような映像検索システムは、顔特徴量だけを条件とした全件検索機能だけでは不十分である。服特徴量や年齢、性別といった他のメタデータや、カメラや時間帯といったコンテキストとの複合条件に対する検索も効率的に実行する必要がある。

本論文では、全件を対象とした検索とともに、あるイベントが起きた時に直近の映像に対する検索という典型的な2つのユースケースに対応したメタデータ管理技術の実現を目標とする。

3. 設計と実装

3.1 設計目標

映像検索システム向けのメタデータストアの設計目標をまとめる。

本システムは、大規模なカメラシステムをサポートし、映像から抽出されたメタデータを長期にわたって蓄積する。まず、カメラシステムの規模に従い、単位時間あたりに生成されるメタデータサイズの量が増加する。このため、カメラシステムの規模に応じて、メタデータ挿入のスループットを達成できる構成であることが望ましい。長期にわたってメタデータが蓄積されるため、蓄積されたデータサイズに従って、システムが拡張できることが望ましい。

検索に関しては、以下のユースケースを考慮した。一つめは、あるカメラで異常行動を検出した際に、直近の映像から

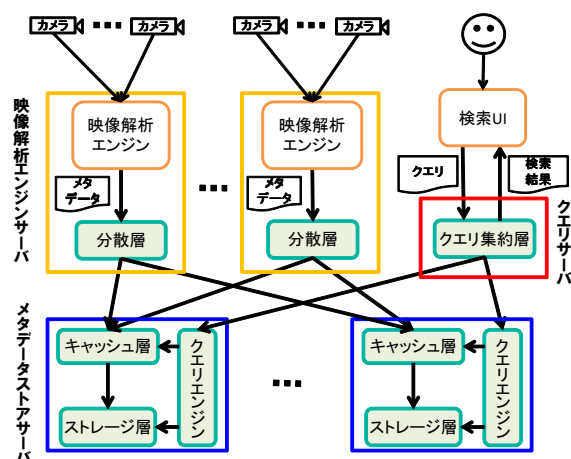


図1 映像検索システム構成

関連人物を探すユースケースである。検索結果を元に、警備員を派遣するなど迅速な行動を取る必要があるため、抽出されたメタデータが即座に検索対象になること、短い応答性能で検索結果を返す必要がある。

二つめは、事件発生後に、メタデータストアに蓄積された全件を対象に目的の人物を探すユースケースである。検索対象となるメタデータサイズはカメラシステムの規模、蓄積された期間に比例するため、それを考慮したデータ配置、検索戦略が必要である。

3.2 映像検索システムの構成

本映像検索システムの構成を図1に示す。映像検索システムは、大きく映像解析エンジンサーバ、メタデータストアサーバ、クエリサーバから構成される。映像解析エンジンサーバはカメラからの映像を解析し、メタデータを抽出する。メタデータストアサーバは、上記で抽出されたメタデータを格納する。映像解析エンジンサーバ、メタデータストアサーバは、カメラシステム規模にあわせて映像検索システムをスケールさせる必要があることから、スケールアウト可能な構成になっている。クエリサーバは、ユーザからの問い合わせを、メタデータストアサーバ群に展開し、その結果を集約してユーザに返す。

以下では、特にメタデータの管理、検索に関わるメタデータストアに関して説明する。

3.2.1 メタデータストア Wally の構成

メタデータストア Wally は、分散層、キャッシュ層、ストレージ層、クエリ集約層、クエリエンジンの5つの機能群から構成される。

分散層は、映像解析エンジンにより抽出されたメタデータをメタデータストアサーバ群に分散させる機能を提供する。このとき、メタデータはカメラの識別子と時刻に基づいて格納先を決定する。このようにメタデータを分散させることで、3.1節で述べた全件検索のユースケースのように、検索対象のカメラ台数が多かったり、長い時間範囲が指定された時、

なるべく多くのメタデータストアサーバ上で検索処理を分散実行が可能となる。

メタデータストアサーバでは、メタデータをキャッシュ層、ストレージ層の二つの階層に格納する。キャッシュ層は直近のメタデータを優先して保持する。これにより 3.1 節で述べた直近映像の検索のユースケースの要件を満たすことができる。ストレージ層はすべてのメタデータを格納する。

メタデータに対する検索は、クエリサーバのクエリ集約層とクエリエンジンが協調することで実現する。クエリ集約層は問い合わせの条件に従い、該当するメタデータが格納されているメタデータストアサーバへ問い合わせを展開する。問い合わせを受けたクエリエンジンは、キャッシュ層、ストレージ層を検索し、条件にあうメタデータを含むフレームを見つける。クエリ集約層は、各クエリエンジンからの検索結果を集約し、その結果を返す。

4. メタデータ検索の具体例

本映像検索システムは、様々な検索機能を提供する。スクリーンショット例を交えて、そのいくつかを紹介する。

4.1 直近の映像検索

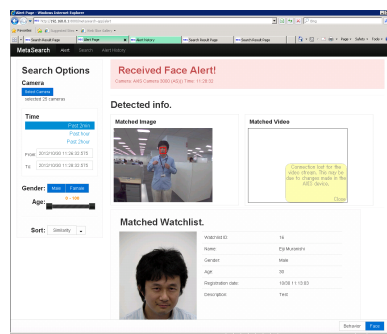
図 2 では、直近の映像検索の一例として、事前に登録された人物の顔を検出し、その時その人物が着ていた服の特徴を用いて直近の映像を探す。つまり、鮮明な顔は得られていないかもしれないが、それらしい服を着た人物が映った映像を見つけたいというシナリオを表現したものである。本システムは、事前に顔画像を登録しておく、その顔画像にヒットする人物が映像に映った時にアラートをあげることができる（図 2(a)）。この時、この映像に映った人物の顔特徴量および服特徴量を解析により抽出できているため、これらの特徴量を元に直近の映像を検索することができる。この時得られた服特徴量を元に検索した結果を図 2(b) に示す。

4.2 オフライン検索

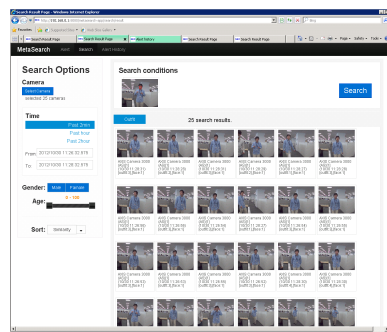
本システムは、様々な条件を指定して、蓄積された映像の検索が可能である。図 3(a) で示す画面を通じて、カメラ集合や時間範囲、顔写真、服の特徴（画像からの抽出、テキスト入力）、推定年齢、性別などの条件を指定することができる。3.2.1 節で述べたように、例えば、時間範囲を広く指定すると、それに比例した数のメタデータストアサーバ群で検索処理が実行される。例えば、服の特徴として、“yellow shirt”というテキストを指定すると図 3(b) のように黄色い服を着た人物が映った様々なシーンが結果として得られる。

5. システムの性能評価

メタデータストアの単体性能を評価した結果を報告する。メタデータストアに格納するメタデータの格納件数を増加させ、メタデータがキャッシュ層にある場合、ストレージ層にある場合の顔特徴量検索、服特徴量検索した際の応答時間を

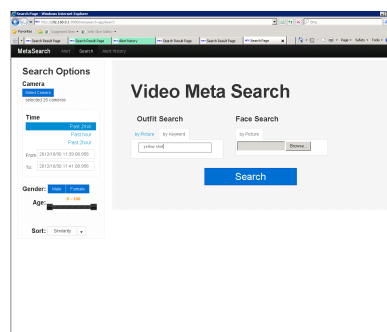


(a) 事前に登録された顔を検出

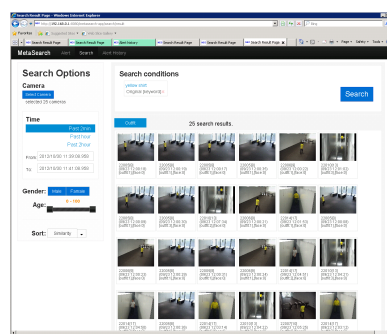


(b) 着ていた服特徴量での検索結果

図 2 直近の映像検索の例（顔を検出し、服で検索）



(a) オフライン検索画面．“yellow shirt”を指定



(b) オフライン検索結果

図 3 オフライン検索の例（テキストによる服特徴量検索）

測定した。その結果を図 4 にまとめる。

どの結果とも、メタデータの量に比例した実行時間がかかっている。キャッシュ層、ストレージ層にあるメタデータ

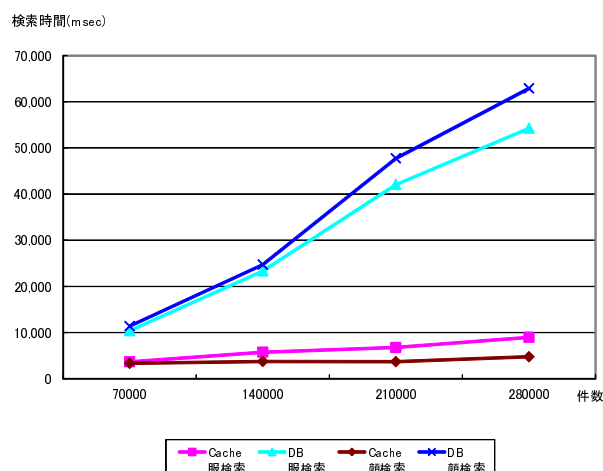


図 4 検索時間（単体性能）

に対する検索実行時間を比較すると、キャッシュ層にある方が、ストレージ層にあるより 10 倍程度高速である。この結果より、映像検索システムのワークロードを考慮し、どのメタデータをキャッシュ層に保持するかが重要であることがわかる。

6. 考察・まとめ

本論文では、映像検索システム向けのメタデータストアを紹介した。大規模なカメラシステム、長期間に渡るメタデータを蓄積する必要から、スケールアウト可能なデータストアを開発した。さらに、典型的な映像検索のユースケースに対応して効率的な検索性能を達成するためのデータ配置アルゴリズムも開発した。そして、実用的なシナリオに対して本システムが機能することをスクリーンショットを交えて示した。

今後の課題として、まず性能評価を充実させ、本データストアのスケール性を確認したい。また、単体性能の評価結果より、メタデータがキャッシュ層にある場合に比べ、ストレージ層にある場合の性能劣化が大きいことがわかった。この劣化を抑えるために、検索効率を向上させるより高度なデータ構造の研究が必要である。

文 献

- [1] 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, Vol. 52, No. 1, pp. 71–78, Jan. 2011.
- [2] Neoface. <http://www.nec.co.jp/soft/neoface/product/neoface.html>.
- [3] Fieldanalyst 性別・年齢層自動推定システム. http://www.nec.co.jp/solution/d_signage/fieldanalyst/index.html.
- [4] 小山和也, 有熊威, 白石展久, 永井洋一, 河又恒久. 大規模映像解析システム向けの解析制御ミドルウェアの提案. 第 11 回情報科学技術フォーラム (FIT2012) 講演論文集, 第 3 巻, pp. 115–116, 2012.
- [5] 有熊威, 小山和也, 白石展久, 永井洋一, 河又恒久. 大規模映像解析システム向けの解析制御ミドルウェアの試作. 第 11 回情報科学技術フォーラム (FIT2012) 講演論文集, 第 3 巻, pp.

117–118, 2012.

- [6] 前川敬介, 内海ゆづ子, 岩村雅一, 黄瀬浩一. 100 万顔画像データベースに対する 34ms での照合の実現. 信学技報, Vol. 111, pp. 95–100, Dec. 2011.