

結果の多様化と要約化に基づいた映像検索システム

董 テイ テイ[†] 西村 祥治[†] 劉 健全[†]

[†] NEC システム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: †t-dong@ce.jp.nec.com, †s-nishimura@bk.jp.nec.com, ††j-liu@ct.jp.nec.com

あらまし 本論文では、検索結果を自動的に多様化して要約する映像検索システムを紹介する。映像検索において、ユーザはしばしば条件指定におけるジレンマに直面する。条件を多く指定するほど、検索結果が少なく不足になる。一方、指定条件の数が少ないと、検索結果が膨大で重複かつ冗長になる。特に、映像の連続フレーム間は重複性が高く、冗長な検索結果につながりやすい。本システムでは、重複または類似している結果を要約し、多様化した代表的な結果を提示する。それにより、条件指定におけるジレンマを解消し、効率的な映像検索を実現する。実験では、監視カメラからの実映像データを用い、本手法の有効性を示す。

キーワード 映像検索, 多様化, 要約

1. はじめに

インターネットや監視カメラなどからの映像データの増加に合わせて、大量の映像データに対する検索技術が急速に発展している。撮影時間場所や、内容を記述したタグなどといったメタデータだけでなく、画像解析技術の進展により、人物の顔や年齢、性別、服色から、一般物体の車や靴まで認識でき、様々な情報が検索できるようになった。ユーザはキーワードを入力したり、例示画像を与え条件を指定したりして自由に検索できる [6]。しかしながら、条件が不適切または間違いがあるなどにより、ユーザはしばしば条件指定におけるジレンマに直面する。例えば、条件を多数指定するほど、検索結果が少なく不足になるのに対して、指定条件の数が少ないと、検索結果が膨大で重複かつ冗長になる。さらに、映像の場合、連続フレーム間は重複性が高く、冗長な検索結果につながりやすい。

図 1 に示している例を用いて説明する。この例では、ユーザは映像データからの画像 {1, 2, 3, 4} に対して、例示画像 q を入力し、 q と類似している上位三件の画像を検索する。既存の類似度に基づいた検索処理手法は画像間の類似度だけを計算する。このため、顔または服色のどちらが類似していると指定する場合、結果として {1, 2, 3} (顔) または {1, 2, 4} (服色) が返され、いずれも重複している画像 {1, 2} が含まれる。一方、顔と服色両方を指定する場合、結果が {1, 2} となる。映像データが大量になるほど、重複している結果が多くなり、ユーザが検索結果を確認するには無駄な時間がかかる。また、先行研究 [7-9] でもこのような問題が起り、同じ人の顔が何度も結果に現れる。それらを受けて、われわれは検索結果を自動的に多様化して要約する映像検索システムを提案する。本システムでは、重複または類似している結果を要約し、多様化した代表的な結果を提示する。それにより、条件指定におけるジレンマも解消でき、効率的な映像検索を実現する。図 1 の例の場合、本システムは画像 2 を画像 1 に要約し、多様化した結果 {1, 3, 4} を返す。類似する結果がまとまっているため、ユーザにとっては直観的で理解しやすく、効率的に結果を確認することができる。

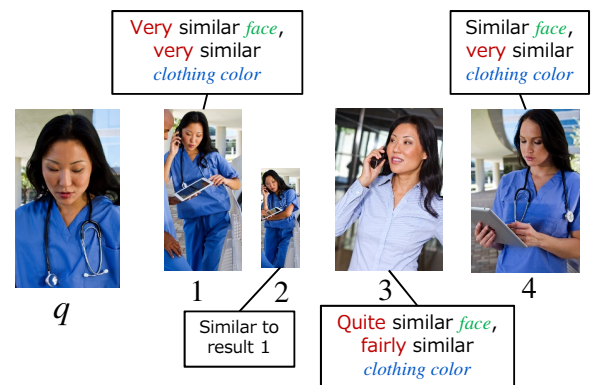


図 1 画像 {1, 2, 3, 4} から、例示画像 q と類似している上位三件の画像を検索する。われわれの方法では、画像 2 を画像 1 に要約し、多様化した結果 {1, 3, 4} を返す。画像 3 は服色、画像 4 は顔において多様化する。画像 2 は顔も服色においても画像 1 と類似しているため、画像 1 に要約される。

検索結果の多様化において、情報検索 (IR) やデータベース (DB)、マルチメディア (MM) の分野では、様々な手法が提案されてきた [1-3, 5, 10, 12, 17-20, 22]。既存研究では、主に単一種類のデータが対象であり、特有の性質に基づくため、他の種類のデータと複数種類の情報が混在するマルチメディアデータに適用できないという問題がある。また、クラスタリングや機械学習を用い、画像データの検索結果を多様化する方法もある [21]。しかし、これらの方法では、ユーザが検索条件やパラメータを調整する際は直観的ではなく、利用しにくい。

本システムは、先行研究 [4] に提案された、複数の特徴量を用いた多様化と要約の手法を利用する。監視カメラからの実映像データを用い、顔と服色に加え年齢と性別を抽出し、検索結果を多様化して要約する映像検索システムを実現した。以下では、文献 [4] の手法を紹介し、システムの実現について説明する。

2. 複数の特徴量を用いた多様化と要約

文献 [4] では、映像などのマルチメディアデータから抽出された複数の特徴量に対し、特徴量の表現力と独立力を定義する。

複数の特徴量を有する各オブジェクト o_i の多様性スコアを計算し、検索結果の多様化と要約を行う。

各特徴量 f_m はそれぞれの情報抽出方法または認識エンジンにより抽出される。特徴量 f_m において、オブジェクト o_i と o_j の類似度は $Sim_{o_i, o_j}(f_m)$ で表される。類似度の計算はブラックボックスとして、コサイン (cosine) など公知の類似度定義を使用する。すべての特徴量における類似度 Sim_{o_i, o_j} は各特徴量における類似度の総和で定義される。

オブジェクトの多様性は各特徴量の以下の性質 (1) オブジェクトを表現または特定できる各特徴量の重要度と (2) 特徴量間の相関により、他の特徴量から各特徴量を推定できない可能性に依存すると考えられる。この考えをもとに、特徴量の表現力 (RP: Representation Power) を、この特徴量でオブジェクトを表現する可能性で定義し、特徴量の独立力 (IP: Independence Power) を、この特徴量は他の特徴量に相関されないまたは独立する可能性で定義する。

オブジェクト o_i の特徴量 f_m の表現力は $RP(f_{m,i})$ で表され、

$$RP(f_{m,i}) = \frac{1}{\|R = \{o_p | Sim_{o_i, o_p}(f_m) \geq \tau(f_m), o_p \in C\}\|}$$

で計算される。 $\|R\|$ は集合 R の基数である。 $\tau(f_m)$ は f_m における類似度閾値であり、ユーザにより指定されるもしくはデフォルト値として設定される。 C はデータセットから得られた候補オブジェクトの集合である。

オブジェクト o_j の特徴量 f_b に対して、オブジェクト o_i の特徴量 f_a の独立力は $IP(f_{a,i}; f_{b,j})$ で表され、

$$IP(f_{a,i}; f_{b,j}) = \frac{\|R_a = \{o_p | Sim_{o_i, o_p}(f_a) < \tau(f_a), o_p \in R_b\}\|}{\|R_b = \{o_p | Sim_{o_p, o_j}(f_b) \geq \tau(f_b), o_p \in C\}\|}$$

で計算される。特に、同じ特徴量 f_m における独立力は非類似度で定義される： $IP(f_{m,i}; f_{m,j}) = 1 - Sim_{o_i, o_j}(f_m)$ 。複数の特徴量に関する独立力は

$$\begin{aligned} IP(f_{m,i}; f_j) &= IP(f_{m,i}; f_{1,j}, \dots, f_{d,j}) \\ &= IP(f_{m,i}; f_{1,j}) \times \dots \times IP(f_{m,i}; f_{d,j}) \end{aligned}$$

で定義される。

オブジェクト o_j に対して、オブジェクト o_i の多様性スコア $DIV(o_i; o_j)$ は各特徴量の表現力と独立力から計算される：

$$DIV(o_i; o_j) = \sum_{m=1}^d RP(f_{m,i}) \times IP(f_{m,i}; f_j).$$

多様性スコアが大きいほど、オブジェクトが多様である。検索結果の多様化と要約では、多様性スコアが最も高いトップオブジェクトを選択し、他の候補をトップオブジェクトに割り当てる。各候補は最も類似しているオブジェクトに割り当てられる。

3. システムの実現

3.1 データセットと特徴量抽出

本システムは監視カメラからの実映像データを用いる。映像データは三日間にわたって、三つの場所 (コンビニストア、電気店、銀行 ATM) で撮影された。各場所では、三つのカメラが



図 2 映像データの例

図 3 検索条件の入力

配置され、色々なシーンに合わせて、複数の映像を撮影した。各映像の撮影時間は 5~10 分である。これらの映像データから、合計約 40 分を選択し、データセットを作成した。図 2 に映像データの例を示している。映像のフレームを抽出し、11279 枚の画像データに変換した。その上、顔や服色、年齢、性別の特徴量抽出を行った。

顔の特徴量について、商用製品 NeoFace[®](注1) を用いて抽出した。人物の年齢と性別も抽出した [15, 16]。Darknet フレームワーク [13] のオブジェクト検出技術 YOLO [14] を用い人物の矩形を検出し、服色の HSV 特徴量を抽出した。抽出された各特徴量と撮影時間場所を統合し、人物の特徴量データベース (16373 エントリ) を作成した。

3.2 システムの画面と検索例

「薄緑色の服を着ている 60 歳以下の女性」という検索例を用い、システムを説明する。図 3 に検索条件「服色：薄緑、年齢：[0, 60]、性別：女性」の入力画面を示している。服色はカラーピッカーのほか、画像をアップロードすることでも指定できる。顔の場合、画像をアップロードして指定する。また、撮影時間場所による絞り込みもできる。

本システムでは、既存の類似度による検索処理手法と [4] に提案された手法を実現し、それぞれの検索結果を比較する。図 4 と図 5 は検索例「薄緑色の服を着ている 60 歳以下の女性」に対して、既存手法と提案手法の処理結果を示している。今回は、索引を使わずに特徴量データベース全件をスキャンし、入力条件に満たしている候補 120 件を抽出した。既存手法では、抽出された候補をそのまま結果として返す。提案手法は抽出された候補に対し、多様化と要約を行い、上位 5 件の結果を返す。

(注1) : NeoFace[®] は米国政府機関の評価テストにおいて [11] 世界 No.1 の精度・速度を達成している顔認証技術である。

Existing method

Processing time: 0.720131 seconds



図 4 既存手法の検索結果

Proposed method

Processing time: 1.337106 seconds



図 5 提案手法の検索結果

処理時間において、既存手法は約 0.72 秒、提案手法は約 1.34 秒がかかった。既存手法に比べ、提案手法では処理時間がより長い、雑然とした結果を類似するもの同士で要約されているので、確認時間を節約することができる。画像一枚の確認時間を 0.2 秒とすると、検索結果を確認するために、既存手法では 24 秒、提案手法では 1 秒がかかる。検索処理時間と確認時間を合計した全体時間で両手法を比較すると、既存手法 (24.72 秒) よりも提案手法 (2.34 秒) の方が 10 倍以上速いことがわかる。

4. まとめと今後の課題

本論文では、監視カメラからの実映像データを用い、顔と服色に加え年齢と性別を抽出し、結果の多様化と要約化に基づいた映像検索システムを紹介した。処理時間において、提案手法は既存手法より遅いだが、結果の確認時間も含めると、提案手法は既存手法より 10 倍以上速いことが分かった。今後は、特徴量に対する索引技術を用い、候補抽出時間と検索処理時間を短縮し、大規模映像データに対するスケーラビリティを高める。

文 献

- [1] Rakesh Agrawal, Sreenivas Gollapudi, Alan Halverson, and Samuel Ieong. Diversifying search results. In *WSDM*, pages 5–14, 2009.
- [2] Gabriele Capannini, Franco Maria Nardini, Raffaele Perego, and Fabrizio Silvestri. Efficient diversification of web search results. *PVLDB*, 4(7):451–459, 2011.
- [3] Van Dang and W. Bruce Croft. Diversity by proportionality: an election-based approach to search result diversification. In *ACM SIGIR*, pages 65–74, 2012.
- [4] Tingting Dong, Shoji Nishimura, and Jianquan Liu. Diversified and summarized video search system. In *ACM MM*, pages 1263–1264, 2017.
- [5] Marina Drosou and Evaggelia Pitoura. Multiple radii disc diversity: Result diversification based on dissimilarity and coverage. *ACM Trans. Database Syst.*, 40(1):4, 2015.
- [6] Jianquan Liu, Shoji Nishimura, and Takuya Araki. Wally: A scalable distributed automated video surveillance system with rich search functionalities. In *ACM MM*, pages 729–730, 2014.
- [7] Jianquan Liu, Shoji Nishimura, and Takuya Araki. Antilooter: A loitering discovery system for longtime videos across multiple surveillance cameras. In *ACM MM*, pages 675–679, 2016.
- [8] Jianquan Liu, Shoji Nishimura, and Takuya Araki. Visloiter: a system to visualize loiterers discovered from surveillance videos. In *SIGGRAPH*, pages 47:1–47:2, 2016.
- [9] Jianquan Liu, Shoji Nishimura, Takuya Araki, and Yuichi Nakamura. A loitering discovery system using efficient similarity search based on similarity hierarchy. *IEICE Transactions*, 100-A(2):367–375, 2017.
- [10] Ziyang Liu and Yi Chen. Differentiating search results on structured data. *ACM Trans. Database Syst.*, 37(1):4, 2012.
- [11] NIST. Multiple Biometrics Evaluation, 2010. www.nist.gov/itl/iad/ig/mbe.cfm.
- [12] Davood Rafiei, Krishna Bharat, and Anand Shukla. Diversifying web search results. In *WWW*, pages 781–790, 2010.
- [13] Joseph Redmon. Darknet: Open source neural networks in c. <http://pjreddie.com/darknet/>, 2013–2016.
- [14] Joseph Redmon and Ali Farhadi. Yolo9000: Better, faster, stronger. *arXiv preprint arXiv:1612.08242*, 2016.
- [15] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Dex: Deep expectation of apparent age from a single image. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, December 2015.
- [16] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision (IJCV)*, July 2016.
- [17] Rodrygo L. T. Santos, Craig Macdonald, and Iadh Ounis. Exploiting query reformulations for web search result diversification. In *WWW*, pages 881–890, 2010.
- [18] Rodrygo L. T. Santos, Craig MacDonald, and Iadh Ounis. Search result diversification. *Foundations and Trends in Information Retrieval*, 9(1):1–90, 2015.
- [19] Reinier H. van Leuken, Lluís Garcia Pueyo, Ximena Olivares, and Roelof van Zwol. Visual diversification of image search results. In *WWW*, pages 341–350, 2009.
- [20] Zhengwei Yang, Ada Wai-Chee Fu, and Ruifeng Liu. Diversified top-k subgraph querying in a large graph. In *SIGMOD*, pages 1167–1182, 2016.
- [21] Maia Zaharieva. An adaptive clustering approach for the diversification of image retrieval results. In *MediaEval*, 2016.
- [22] Kaiping Zheng, Hongzhi Wang, Zhixin Qi, Jianzhong Li, and Hong Gao. A survey of query result diversification. *Knowl. Inf. Syst.*, 51(1):1–36, 2017.