

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 20 日現在

機関番号：10101
 研究種目：基盤研究（B）
 研究期間：2009～2012
 課題番号：21300030
 研究課題名（和文）画像・音響信号およびテキストデータ融合利用による異種メディア横断検索
 研究課題名（英文）Cross-media retrieval through collaborative use of image/audio signals and text data
 研究代表者
 長谷山 美紀（HASEYAMA MIKI）
 北海道大学・大学院情報科学研究科・教授
 研究者番号：00218463

研究成果の概要（和文）：画像や映像、音楽などのメディアが固有に持つユーザの曖昧な要求に応える検索理論の構築との全体構想に基づき、ユーザが映像や画像、音楽をクエリとして与えた場合に、望むコンテンツを推定し、効果的に提示する手法を実現した。具体的には、その実現のために、画像、映像、音楽の 3 つのメディアに検索対象を絞り、メディア横断型検索を実現し、ユーザがキーワードや同一メディアコンテンツのクエリを提示できない場合にも、希望するコンテンツを取得可能とした。

研究成果の概要（英文）：

We implemented an effective method for recommending the desired contents to a user. The method accepts an image, video or music clip as a query, and is based on search theory that can deal with ambiguous user requests that are unique to each image, video or music clip. Specifically, in order to implement this method, we narrow down the search results to the three media types of image, video and audio, and implement cross-media search, allowing the user to acquire the desired contents even in cases where they cannot express the same query using keywords or media of the identical type.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009 年度	4,300,000	1,290,000	5,590,000
2010 年度	4,000,000	1,200,000	5,200,000
2011 年度	2,700,000	810,000	3,510,000
2012 年度	1,400,000	420,000	1,820,000
年度			
総計	12,400,000	3,720,000	16,120,000

研究分野：総合領域

科研費の分科・細目：メディア情報学・データベース

キーワード：情報検索

1. 研究開始当初の背景

映像検索は、一般に映像データに付与されたメタデータを指標として行われる。大量の映像データにアクセスする今日の状態を予測し、自動でメタデータを付与する技術が研究され、以前より様々な手法が提案されてい

た。これら手法の多くは、映像における物理特徴（低レベルメタデータ）に注目したものであり、真にユーザが望む映像の検索のためには、セマンティックメタデータを付与する必要がある。過去において、このような技術の実現は、非常に困難であるとされていたが、

ここ数年、コーパスベースのメディア解析手法が提案され、映像情報の意味解析技術は実用に向けて大きく前進したと言える。

しかしながら、コーパスベースの手法を高度化しても、映像固有の視聴形式から、ユーザが満足する映像検索結果を得るためには、依然として大きな問題が存在する。それは、コーパスベースの手法により実現されるのは、セマンティックメタデータの付与であり、高精度なメタデータ付与が実現されたとしても、ユーザがクエリとなるキーワードを明確に与えることができない場合には、検索結果へ満足感を覚えれないと言うものである。つまり、高精度なメタデータ付与手法は、映像情報が持つ多様性や曖昧さなどに柔軟に対応することを可能とするセマンティックメタデータを付与することはできるものの、その検索方式は、キーワードの提示を要求する検索形態に限定され、ユーザの曖昧な要求に応えることはできない。同様の問題が、映像検索だけでなく、音楽や静止画像の検索にも生じ、これは、WWWにおけるテキスト検索との大きな差異と考えられる。このような差異はプル型の情報サービス（必要な情報を、ユーザが能動的に取得するタイプの情報サービス）と位置づけられているテキスト検索とは対照的に、画像・映像検索や音楽検索には、ユーザの能動的な操作を伴わずに、自動的に情報を配信するプッシュ型情報取得が適する場合があることから知ることもできる。つまり、画像や映像、音楽などのメディアにおいて、メタデータを付与する研究を進めるだけでは、ユーザが望むコンテンツを提示することが困難であり、本研究では、それを可能とする理論構築を目指し、その実現のために必要となる一技術として『メディア横断型検索手法』を提案する。また、研究代表者は、映像において固有の検索方式の必要性を主張し、映像検索インタフェースを提案している。この研究成果を踏まえ、映像に限定されない、新しいメディア横断型の検索手法の提案と言う本申請の着想に至った。

2. 研究の目的

本研究では、理論構築に必要な一手法である、ユーザが映像や画像、音楽などをクエリとして与えた場合に、望む映像や画像、音楽を推定し、効果的に提示可能なメディア横断型検索手法を提案する。本手法により、多様なメディアのコンテンツの視聴履歴から、個人の嗜好を抽出し、異種メディア間での検索を可能とすることで曖昧なユーザの検索要求に応えるための1技術が実現される。従来より、ユーザが画像をクエリとして与えた場合に、同一、もしくは、類似画像を提示する手法や、映像から映像、音楽から音楽を検索する手法（以降、同一メディア内検索と呼ぶ）は、各種提案されてきたが、異種メディアを

横断する検索手法は、提案されておらず、独創的かつ先駆的と言える。

3. 研究の方法

本研究の全体構想は、画像および映像、音楽メディアに固有のユーザの曖昧な要求に応える検索を実現する理論構築にあり、本研究は、その理論構築に必要な一手法である、ユーザが映像や画像、音楽などをクエリとして与えた場合に、ユーザが望む映像や画像、音楽を推定し、効果的に提示可能な異種メディア横断型検索手法の提案にある。その実現のために、画像、映像、音楽の3つのメディアに検索対象を絞り、次の方法で手法の構築を目指す。

(1) (実現ステップ1) 異種メディアコンテンツの距離尺度の定義

個人の視聴コンテンツの集合から抽出される物理特徴量と、コンテンツの視聴履歴から、同一メディアおよび異なるメディアの二つのコンテンツ間の距離を定義する。

(2) (実現ステップ2) 個人の嗜好を表現可能なモデルの実現

(実現ステップ1)で定義された異種メディアコンテンツ間の距離尺度を用いることで個人の好み表現可能なモデルを実現する。

(3) (実現ステップ3) 嗜好モデルの類似性の可視化システムの実現

(実現ステップ2)で実現したモデルを用いて異なるユーザの嗜好モデルの類似性を可視化のために、インターネットにおけるネットワークモデル利用について検討する。ユーザ嗜好モデルの類似性の可視化システムをソフトウェアにて実現する。また、本研究で提案した各種手法について公開に向けたソフトウェアの整備を行う。

4. 研究成果

上記3. 研究の方法(1)~(3)で記載した各ステップについてそれぞれの成果を下に記載する。

(1) (実現ステップ1) 異種メディアコンテンツの距離尺度の定義について

提案の距離尺度は各コンテンツから、色特徴に注目した有向グラフを生成し、そのカット辺を比較することで定義する。ここで、提案される定義の距離尺度を算出するためには、映像コンテンツをシーン分割する必要がある。複数の閾値をあらかじめ設定する必要がある。閾値を変動させることでシーン分割の精度は大きく変化する。しかしながら、シーン分割結果を安定して求めなければ、信頼できる距離を算出することはできない。それらの最適な値を見つけ出し、設定することは容易ではないが、我々は、これらの閾値を変更してシーン分割を複数回行い、得られた複数の結果を統合することで最終的な結果を取得する手法を見出した。これにより、提案手法では事前の閾値設定が不要となる。詳細

を下に説明する。

今、 n 個のコンテンツを $I_i^{(n)}$ ($i=1,2,L,N_s-1$) とし、各コンテンツのシーン s_i および s_i+1 間にカット辺が存在する場合に、 $I_i^{(n)}=1$ 、存在しない場合には $I_i^{(n)}=0$ とする。すなわち、 $I_i^{(n)}$ は次式のように定義される。

$$I_i^{(n)} = \begin{cases} 1 & \text{if a cut edge exists between } s_i \text{ and } s_{i+1} \text{ in } n\text{-th result} \\ 0 & \text{otherwise} \end{cases}$$

ここで、提案手法では 2 種類の有向グラフを生成しているため、各有向グラフにおいて閾値 $Th_{T^{vp}}$ と $Th_{C^{vp}}$ の値を変更することで得られるシーン分割結果 $I_i^{(n^{vp})}$ ($n^{vp}=1,2,L,N^{vp}$) および各シーンの類似性を判定する閾値 Th_{T^A} と Th_{C^A} の値を変更することで得られるシーン分割結果 $I_i^{(n^A)}$ ($n^A=1,2,L,N^A$) の 2 種類の結果が得られる。そこで、提案手法ではこれらの結果をそれぞれ以下の式で統合し、 s_i および s_{i+1} 間にシーン境界が存在する確率 P_i^{vp} および P_i^A を算出する。

$$P_i^{vp} = \frac{1}{N^{vp}} \sum_{n^{vp}=1}^{N^{vp}} I_i^{(n^{vp})}$$

$$P_i^A = \frac{1}{N^A} \sum_{n^A=1}^{N^A} I_i^{(n^A)}$$

最後に、提案手法では、最終的なシーン分割結果を以下の式で取得する。

$$P_i = P_i^{vp} \cdot P_i^A$$

最終的に、対象映像において、 P_i が閾値 T_i よりも高い値を示す協会シーン境界として分割する。

以上により、提案手法では、映像中の色特徴と出演者情報および音響特徴に注目して得られるそれぞれのシーン分割結果を統合することで、色画像特徴と音響特徴の両者を利用したコンテンツの差異を算出することが可能となった。

(2) (実現ステップ2) 個人の嗜好を表現可能なモデルの実現について

従来より、映像等のコンテンツを推薦する手法として、協調フィルタリングに基づく手法が種々提案されている。協調フィルタリングでは、各ユーザの映像に対する評価値の類似度より、推薦対象ユーザと好みが類似するユーザ (以降、類似ユーザ) を選択する。そして、類似ユーザの評価値に基づき、推薦対象映像群に対する評価値を推定することで映像の推薦を行う。しかしながら、協調フィルタリングに基づく手法では、推薦対象映像群の申で評価値を付与された映像が少ない場合、推薦の精度が低下する問題が存在する。また、新しく推薦対象映像群に追加される映像は、評価値を付与されない限り推薦結果として提示されないという問題も指摘されている。

一方で近年、映像を視聴するユーザから得られる特徴より、映像に対する関心の度合い (以降、関心度) を推定し、これに基づき映像の推薦や要約を行うことを目的とした研究が進められている。関心度はユーザの現在の好みを表すため、これを用いることでユーザの好みの変化に対応した高精度な映像の推薦が可能となると考えられる。その内の 1 つでは、映像視聴時におけるユーザの顔から取得される特徴量を用いて、Hidden Markov Model (HMM) に基づき関心度を推定している。また、他の手法では、表情、頭部の動き、瞬目動作、視線より得られる特徴量を用いて、ファジィ推論に基づき関心度を推定している。これらの手法では、映像視聴時のユーザの顔から取得される特徴量のみを用いて関心度を推定している。一方さらに異なる論文では、映像視聴時におけるユーザの瞬目動作、視線変動、身体動作 (以降、視聴動作) とユーザが視聴する映像 (以降、視聴映像) より得られる特徴量を用いて、サポートベクトル回帰に基づき関心度を推定している。このように、視聴動作と視聴映像より得られる特徴量を用いることで、ユーザの顔が検出できない場合でも関心度を推定することが可能となる。しかしながら、これらの手法では、関心度を推定する際に類似ユーザを考慮していない。一般に、類似ユーザを考慮することで推薦対象ユーザの好む未知の映像の推薦が期待される。そのため、類似ユーザを考慮して関心度を推定することで、高精度な映像の推薦が可能となると考えられる。そこで、本事業では、視聴動作と視聴映像より得られる特徴量から、類似ユーザを考慮して関心度を推定し、これを用いて映像の推薦を行う手法を提案した。提案手法では、まず、推薦対象ユーザより得られる視聴動作と視聴映像に関する特徴量を用いて、Support Vector Machine (SVM) に基づき関心度を推定する。このとき、類似ユーザの学習データを用いることで、これを考慮した関心度の推定を可能とする。ここで、視聴映像の関心度が最大となる時間における映像特徴量と、推薦対象映像群より得られる映像特徴量との類似度を用いて映像の推薦を行った。しかしながら、推薦対象ユーザは関心度が最大となる映像と類似する映像にのみ関心を示すとは限らない。そのため、関心度と視聴映像との関係性に基づき、推薦対象映像群に対する関心度を推定することが必要となる。本事業では、類似ユーザを考慮して推定される関心度を視聴映像との関係性を考慮して、映像の推薦を行う手法を提案する。提案手法では、まず、先の手法と同様に、類似ユーザを考慮して推薦対象ユーザの視聴映像に対する関心度を推定する。次に、得られた関心度と視聴映像との関係性を考慮するために、これらを用い

てサポートベクトル回帰の学習を行う。そして、得られたモデルに基づき推薦対象映像群に対する関心度を推定することで、映像の推薦を行う。これにより、類似ユーザを考慮して推定される関心度と視聴映像との関係性を考慮した、高精度な映像の推薦が可能となった。

(3) (実現ステップ3) 嗜好モデルの類似性の可視化システムの実現について

本研究助成の事業では画像間の類似度に基づく可視化による画像検索結果の提示法の有効性の走査評価のために、人間の視覚特性を導入した評価指標を提案した。提案した評価指標は、視覚特性に関する2種類の評価値を組み合わせることで得られる。まず、人間の視覚特性を表す代表的なモデルである特徴統合理論を用いた評価値を算出する。特徴統合理論において、人間は視界において探索の対象物体が存在する領域が視覚的に顕著である程、容易に見えてくるといった性質を持つことが示されている。これより、特徴統合理論に基づいて提示された画像検索結果における視覚的顕著性 (visual saliency) を、ユーザが画像検索結果から所望の画像を探索するときの提示法の有効性を表す評価値として求める。さらに提案手法では、視覚的顕著性の算出は顕著性マップを用いることで実現した。次に、ユーザは類似した画像が互いに近接し、まとまり (以降、群と呼ぶ) を形成している程、各々の画像の探索を容易に行えるという性質に基づき、提示された画像検索結果において類似した画像が群をなす度合いを、提示法の有効性に関する指標として算出する。最後に、得られたこれらの評価値を組み合わせることで、提示法の評価指標を導出している。提案した評価指標により、画像検索結果の提示法が有効性を示す場合について、視覚特性に基づいて定量的に評価することを可能とした。

(4) 上記成果をふまえ、提案手法を含む映像コンテンツ推薦システムの実証実験を2度にわたり実施した。以下に、その日時他を記す。

①第1回実証実験

期間：2012年9月11日～30日

場所：札幌駅前地下歩行空間

本実証実験について下の報道が行われた。

- ・北海道新聞9月12日朝刊27面
- ・札幌テレビ、北海道テレビなど

②第2回実証実験

期間：2013年2月2日～11日

場所：札幌駅前地下歩行空間

本実証実験について下の報道が行われた。

- ・北海道新聞2月2日朝刊25面
- ・テレビ北海道など

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計25件)

① Miki Haseyama, Takahiro Ogawa, Trial Realization of Human-Centered Multimedia Navigation for Video Retrieval, International Journal of Human-Computer Interaction, 査読有, vol. 29, pp. 96-109, 2013

② Miki Haseyama, Daisuke Izumi, Makoto Takizawa, Super-Resolution Reconstruction for Spatio-Temporal Resolution Enhancement of Video Sequences, IEICE Transactions on Information and Systems, 査読有, E95-D, 2012, pp. 2355-2358

③ Takahiro Ogawa, Miki Haseyama, Adaptive Single Image Superresolution Approach Using Support Vector Data Description, EURASIP Journal on Advances in Signal Processing, 査読有, 2011

DOI: 10.1155/2011/852934

④ Takahiro Ogawa, Miki Haseyama, Missing Intensity Interpolation Using a Kernel PCA-Based POCS Algorithm and Its Applications, IEEE Transactions on Image Processing, 査読有, vol. 20, no. 2, 2011, pp. 417-432

⑤ Takahiro Ogawa, Miki Haseyama, Missing Texture Reconstruction Method Based on Perceptually Optimized Algorithm, EURASIP Journal on Advances in Signal Processing, 査読有, 2010,

DOI:10.1155/2010/208976

[学会発表] (計75件)

① Shuhei Ohnuki, Sho Takahashi, Takahiro Ogawa, Miki Haseyama, Soccer Video Segmentation Based on Team Tactics Estimation Method, Proceedings of International Workshop on Advanced Image Technology 2013, 2013年1月7日, Nagoya University(Aichi)

② Soh Yoshida, Takahiro Ogawa, Miki Haseyama, Accurate Scene Segmentation Method Based on Video Structures in Music Programs, Proceedings of International Workshop on Advanced Image Technology 2013, 2013年1月7日, Nagoya University(Aichi)

③ Akihiro Takahashi, Takashi Hasegawa, Takahiro Ogawa, Miki Haseyama, Saliency Region Estimation for Scanning Electron Microphotographs of Insects Based on

Support Vector Data Description, The 27th International Technical Conference on Circuits/Systems, Computers and Communications, 2012年07月16日, 札幌コンベンションセンター (北海道)

④ Ryosuke Harakawa, Yasutaka Hatakeyama, Takahiro Ogawa, Miki Haseyama, Web Community Extraction with Weighting Three Kinds of Features for Web Video Retrieval, The 27th International Technical Conference on Circuits/Systems, Computers and Communications, 2012年07月16日, 札幌コンベンションセンター (北海道)

⑤ Marie Katsurai, Takahiro Ogawa, Miki Haseyama, A Cross-Modal Approach for Extracting Semantic Relationships of Concepts from An Image Database, 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2012年3月29日, 国立京都国際会館 (京都府)

⑥ Takahiro Ogawa, Miki Haseyama, Perceptually Optimized Subspace Estimation for Missing Texture Reconstruction, 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2012年3月28日, 国立京都国際会館 (京都府)

⑦ Zaixing He, Takahiro Ogawa, Miki Haseyama, Linear Time Decoding of Real-Field Codes over High Error Rate Channels, ICASSP 2011, 2011年5月24日, Prague Congress Centre (チェコ・プラハ)

⑧ 和泉 大佑, 小川 貴弘, 長谷山 美紀, ぶれにより劣化した動画像の復元および高解像度化に関する一検討 -超解像処理における事前確率分布の推定-, 電子情報通信学会技術報告, 2010年6月22日, 北見工業大学 (北海道)

⑨ 宋 妍, 小川 貴弘, 長谷山 美紀, 映像の構造に注目した MCMC 法によるシーン分割の高精度化に関する検討, 電子情報通信学会技術報告, 2010年6月22日, 北見工業大学 (北海道)

⑩ 大串 裕幸, 小川 貴弘, 長谷山 美紀, カーネル正準相関分析に基づく人間の動きに適した楽曲の推薦に関する一検討, 電子情報通信学会技術報告, 2010年6月21日, 北見工業大学 (北海道)

〔図書〕 (計 0 件)

〔産業財産権〕

○出願状況 (計 3 件)

①

名称: 体験記録装置および方法およびプログラム

発明者: 長谷山 美紀

権利者: 国立大学法人北海道大学

種類: 特許

番号: 発明届受付番号: P2012-091

出願年月日: 2012年10月3日

国内外の別: 国内

②

名称: コンテンツ検索装置およびコンテンツ検索プログラム

発明者: 長谷山 美紀

権利者: 国立大学法人北海道大学

種類: 特許

番号: 特願 PCT/JP2009/069005

出願年月日: 2009年11月6日

国内外の別: 国外

③

名称: 動画検索装置および動画検索プログラム

発明者: 長谷山 美紀

権利者: 国立大学法人北海道大学

種類: 特許

番号: 特願 PCT/JP2009/067652

出願年月日: 2009年10月9日

国内外の別: 国外

○取得状況 (計 0 件)

〔その他〕

ホームページ等

<http://www-lmd.ist.hokudai.ac.jp/>

6. 研究組織

(1) 研究代表者

長谷山 美紀 (HASEYAMA MIKI)

北海道大学・大学院情報科学研究科・教授

研究者番号: 00218463

(2) 研究分担者

小川 貴弘 (OGAWA TAKAHIRO)

北海道大学・大学院情報科学研究科・助教

研究者番号: 20524028

荒木 健治 (ARAKI KENJI)

北海道大学・大学院情報科学研究科・教授

研究者番号: 50202742