

放送番組素材検索システムの試作

藤井真人^{†1} 佐野雅規^{†2} 河合吉彦^{†2} 望月貴裕^{†2}

概要: 放送局では担当部局ごとに映像、画像などの番組素材が長年にわたり保存されているが、それらの多くは未整理のままで、簡単には意図した素材を見つけることが難しい。そこで我々は、番組素材の活用を促進するために、映像解析処理を組み込んだ検索システムを試作した。各担当部局の実態に応じてシステムを柔軟に構成でき、導入、運用をできるだけ容易にすることを旨として開発を進めた。本発表では、この試作システムの概要について報告する。

キーワード: 画像検索, 映像データベース, メタデータ

Development of a Search System for TV Program Material

MAHITO FUJII^{†1} MASANORI SANO^{†2}
YOSHIHIKO KAWAI^{†2} TAKAHIRO MOCHIZUKI^{†2}

Abstract: Although producers of broadcast keep a lot of material for TV program, it is difficult to easily find intended images and videos clip due to unstructuring them. We have developed a search system for TV program material. The system is developed with the aim of being able to flexibly configure it according to the situation of each unit in broadcast station. It is also developed to easily able to be introduced and operated. In this paper, an outline of the system is described.

Keywords: Image retrieval, Video database, Metadata

1. はじめに

放送局にはこれまでに蓄えてきた膨大な映像があるが、その管理方法についてはいまだ模索が続いている。なかでも、番組制作に関わるセクションでは、必要に応じて番組素材となる画像や映像などを個別に保管しているケースがあるが、そのほとんどは未整理のままである。これらのデータを有効に活用できれば番組制作の生産性向上や番組内容の充実が期待できる。そこで我々は今回、VFXを担当するセクションで蓄えている写真素材と、宇宙関連番組を制作しているセクションの動画やコマ撮り写真などの素材を管理するシステムをそれぞれ構築した。これらのシステムは、東日本大震災において集まった膨大な素材映像にメタデータを付与する際の支援システム[1]をベースにしている。本報告では新たに開発したこの素材検索システムの概要について述べる。

2. 一般的な素材管理の実態

放送局では、放送済みの番組映像や重要な素材映像をアーカイブスの形で保存している。しかし、アーカイブスの容量には限りがあり、アーカイブスに保存できない素材は、番組制作担当者が共同で保管しているセクションもある。これは、将来、素材を再利用することを想定して行っていることではあるが、実際に利用されるかどうかは定かでは

ないため、その管理にコストをかけることに躊躇がある。そのため、各自がばらばらに保管していたり、少し進んだセクションでも共有フォルダを決めて各自のデータをそこに保存する形態が多い。しかし、共有フォルダに入れる場合でも以下のような不便さがある。

- 他人の情報はわかりにくい。
- 数量が増すと自らのデータを探すことも困難になる。
- 内容を確認するためには逐次クリックして画像を表示したり、動画を再生する必要がある。
- 階層化されている場合には、それらを辿りながら探していく手間が大きい。

このため、コストはかけられないが、実質的に素材を有効利用できる仕組みが望まれていた。今回、この不便さを改善する目的で、素材を決められたフォルダに入れるだけで、あとは検索に必要な処理が自動的に行われ、一定時間後には検索が可能となるシステムを構築したので報告する。

3. 画像検索技術

提案システムは、番組制作の素材となる画像や映像を再利用可能とするためのものである。映像も画像に分割できるため、基本的には画像を検索する機能を備えたシステムを作れば、十分に有用な素材管理ができると考えた。そこで、提案システムには、画像認識により事前に付与したテキスト情報（タグ情報）による検索と、画像をクエリーと

^{†1} (一財)NHK エンジニアリングシステム

NHK Engineering System, Inc.

^{†2} NHK 放送技術研究所

NHK Science and Technology Research Laboratories

する類似画像検索を組み込んだ。はじめに、これらの手法について簡単に述べる。なお各手法の詳細については各文献[2][3]を参照されたい。

3.1 画像認識によるタグの自動付与

画像や映像に映っている物や情景を表す名前(タグ)を、画像認識により自動で付与する機能を組み込んだ。ここで用いた画像認識は、物体や情景などの被写体を認識するためのものであることから、以後被写体認識と呼ぶ。

図1に本稿で用いた被写体認識の事前処理の概要を示す。本手法では、局所的特徴と大局的特徴を合わせた特徴を、認識のための画像特徴とした。

局所的特徴は、見え方の違いの影響を受けにくい SIFT (Scale-Invariant Feature Transform) 特徴と SURF (Speeded Up Robust Features) 特徴をベースとした。まず、これらの特徴を大量の画像から抽出した後、ある程度粗いカテゴリに分類する。このカテゴリを用いて、各カテゴリに当てはまる SIFT 特徴および SURF 特徴の数を画像全体から積算してヒストグラムを作り、これをその画像の局所的特徴とした。

一方、ある範囲の画像領域から取り出す大局的特徴としては、色モーメント特徴、ハールウェーブレット特徴、ローカルバイナリーパターン特徴を使った。

そして、これらの局所的特徴と大局的特徴を結合したベクトルを特徴ベクトルとして画像毎に生成した。

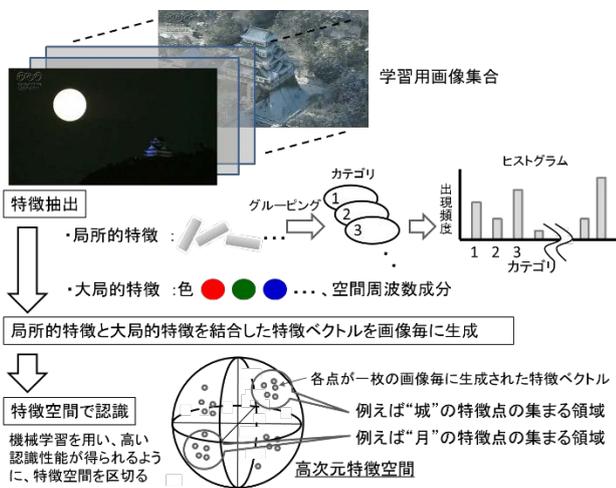


図1 被写体認識処理の概要 (イメージ図)

Figure 1 Process of object recognition.

被写体認識は、特徴ベクトルの持つ次元から構成される特徴空間内の位置に基づいて行う。まず、学習によって、認識したい被写体の画像(正例)と、それ以外の画像(負例)を用い、特徴空間内で対象の被写体の特徴ベクトルが占める領域を決める(図1の領域はあくまでイメージで厳密でない)。今回は、高い認識性能を持つことで知られているサポートベクターマシン(SVM)を用いて学習を行った。学習後に新たな画像が入力されると、その画像から抽

出された同様の特徴ベクトルを特徴空間内にプロットし、その位置から画像が対象の被写体であるか否かを判定する(図2)。実際には、対象の被写体らしさに相当する値をSVMが出力するため、その値を閾値処理することにより判定した。

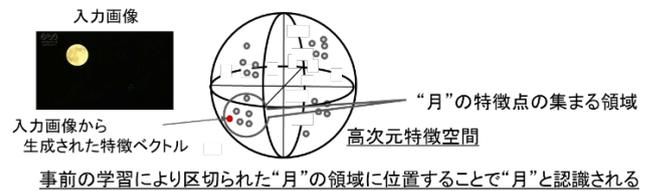


図2 被写体認識 (イメージ図)

Figure 2 Object recognition

なお、本手法では、被写体のサイズに依存せずに対象を認識可能とする工夫も加えている。具体的には、画像から特徴ベクトルを生成する際に、その画像の範囲を、画像全体、各1/2画像領域などと段階的に処理対象とする領域を細かく区分していき、それらの特徴ベクトルを結合したものを最終的な特徴ベクトルとした。

以上に説明した技術を用いると、認識させたい被写体の正例と負例を用いて学習し、学習済みの画像認識モデル(タグモデル)によってあらかじめ検索対象の画像や映像にタグを付与しておくことができるようになる。

3.2 類似画像検索

被写体認識ではあらかじめ認識対象とする被写体を決めて学習する必要がある。そのため、学習されていない被写体の検索はできない。そこで、学習されていない被写体でも、画像をクエリとして画像間の類似性を手掛かりに検索する類似画像検索の機能を組み込んだ。類似画像検索では、たった一枚の画像から似ている物体や情景を探し出すため、画像特徴の選択は重要である。本システムで用いた手法では、色の情報とテクスチャの情報を組み合わせた特徴ベクトルを使った。テクスチャ特徴の抽出法としては、フラクタル的な構造を反映した手法を用いている。

また、構図の類似性も利用したいため、画像をブロックに分割してブロック毎に抽出した特徴ベクトルを比較する手法を使った。但し、均等に分割したブロック同士の間を単純に比較する方法では、被写体の位置の違いによって性能を大きく劣化させる可能性がある。そこで、本手法では、主となる被写体の位置がある程度異なっても似ていると判定できる仕組みを考案した。具体的には、画像中でより目立つ領域を検出し、その領域をなるべくカバーするようにブロックを移動させ、目立つ被写体同士が比較されやすいような工夫を施している(図3)。

以上のような技術を用いることで、クエリ画像と見た目が似ている画像を、その類似度順に表示することなどが可能となる。

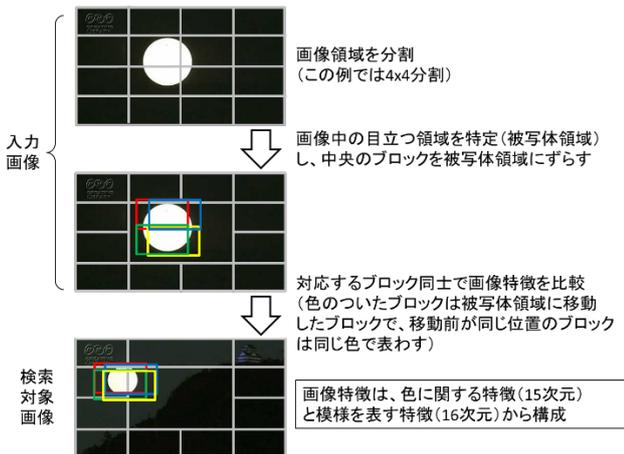


図3 類似画像検索処理の概要
Figure 3 Process of similar image retrieval.

4. 検索システム

上記に述べた二つの検索機能を組み込んだ素材検索システムを試作した。以下にそのシステムについて述べる。

4.1 システムの概要

システムの全体像を図4に示す。本システムを利用するユーザはネットでつながっているコンピュータからブラウザ経由で本システムを使用する。

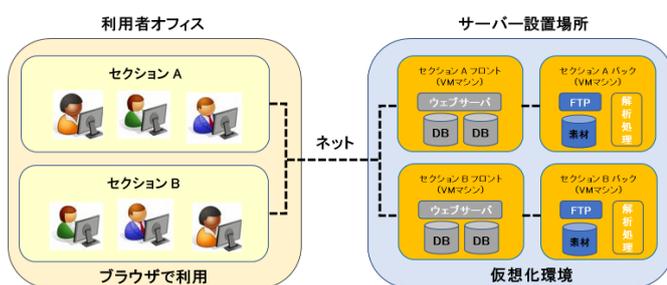


図4 システムの全体像
Figure 4 Overview of the system.

今回は、二つの放送現場セッションに向けてシステムを構築した。二つのセッションでは、検索対象となる素材と検索要求が異なるため、共通部分は多いものの、二つのセッションに向けて別々のシステムを構築した(図4では上の列と下の列がこの二つのシステムに対応)。

図4の右側の薄青で囲んだ四角内が全体で中心となるサーバーシステムで、これを仮想化環境上に構築した。システムはフロントエンドのウェブサーバーと解析処理等を行うバックエンドの2部構成にしている。フロントエンドのDBは、本番系と予備系の2系統で構成した。

静止画から画像特徴を抽出し、検索のためのメタデータを生成するまでのバックエンドの処理は二つのセッション向けのシステムで共通の仕様とした。ここでは、NHK放送技術研究所で提案しているメタデータ制作フレームワーク

(MPF:Metadata Production Framework) [4]に基づいてシステムを構築しており、被写体認識処理後のデータもMPFの規定に則ったXMLデータとして保存した。MPFでは、新たなタグモデルなどもモジュールとして容易に追加可能であり、また、それらを組み合わせた処理も簡単に実現できるなどの利点がある。

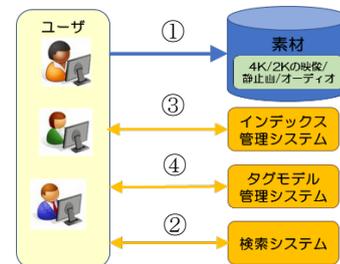


図5 ユーザが操作可能なサブシステム等
Figure 5 Sub-systems for some operations.

ユーザが操作するのは、基本以下の4つで(図5参照)、①決められたフォルダに素材を置く、②タグの自動付与の発動(インデックス管理システムで実行)、③タグモデルの作成(タグモデル管理システムで実行)、④素材検索(検索システムで実行)である。また、管理者権限を与えられたユーザは、新規ユーザの登録や権限の付与などのユーザ管理を行うことができる。図6には、二つのセッションに共通のログイン画面を示した。



図6 ログイン画面
Figure 6 Login screen.

ログイン後は、ユーザ毎に利用できるサブシステムのアイコンが表示される(図7. この場合は、上記②, ③, ④すべてのサブシステムを利用可能なユーザの例)。

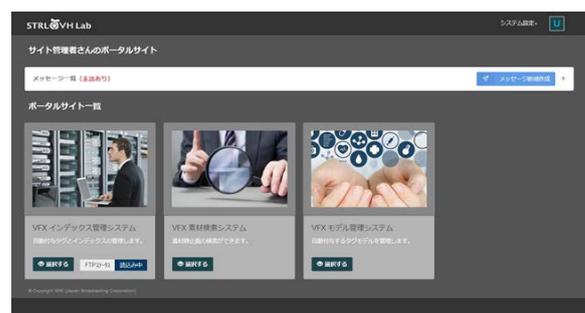


図7 利用可能なアプリケーションのアイコン画面
Figure 7 Display of icons for applications.

以下で②～④のサブシステムについて順次説明する。

4.2 インデックス管理システム

本サブシステムは、基本的にはインデックス作成の開始を指示するためのものである。主には、新たなデータが追加された場合や、4.3 で述べる新たなタグモデルが追加された場合に本サブシステムを使ってインデックスを刷新する。このときインデックス化したいタグのモデルを選択できるようになっている。なお、今回のシステムでは、自動で付与されたタグ情報をあとで手動で修正できるようにしている。そこで、手動で修正された情報が自動付与されたものよりも優先するものとし、以前に手動で与えられた情報はインデックス作成の処理により上書きされたり消されたりすることがないようにしている。

インデックス作成は、2 系統のうち予備系の DB に施され、インデックス作成処理が終了後に自動的に本番系に切り替わる（これまで本番系であったものが予備系となる）。これは、インデックス作成に時間がかかるため、利用時間に制限がかからないようにするための 2 系統化である。また、このような仕組みにしたことで、一代前にはなるが予備系はバックアップの役割も果たしている。

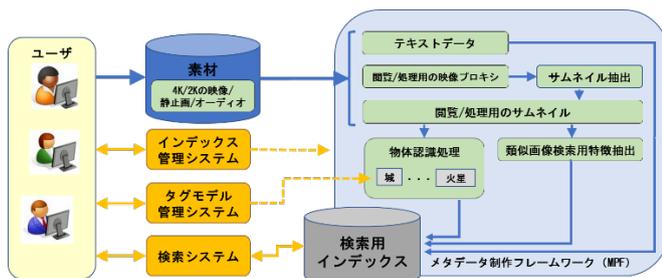


図 8 処理の概略図

Figure 8 Overview of processes of the system

ここで、インデックス作成の指示によって発動される処理について概略を述べておく。これらの処理はバックエンドで行われ、図 8 では水色の大きな四角で囲った部分がこれらに該当する。

本システムでは、ユーザの持つ素材関連情報を取り込むための仕組みを、各ユーザの実情に応じて準備した。VFX のセクションでは、フォルダ名に、番組名－場所－日時という階層化ルールを設けてそのフォルダ名を検索対象とした。宇宙関連番組のセクションでは、番組名や撮影場所、制作担当者名などの情報からも検索したいという要望があり、手動でこれらのテキストデータを入力し、検索に反映できる仕組みを作った。インデックス作成の指示が発動されると、これらの事前で入力された素材関連情報が抽出され、システムに取り込まれる。

同時に、映像解析処理（被写体認識と類似画像検索のための特徴抽出）および閲覧のためのサムネイル画像も生成される。VFX のセクションが所有していた写真画像はサイ

ズもまちまちでフォーマットの種類も 22 に及んだが、すべてからサイズが 240x432 の JPEG のサムネイル画像を生成した。対象が動画の場合には、動画からプロキシ映像が生成され、そこから代表とするサムネイル画像が抽出される。このサムネイル画像も 240x432 の JPEG 画像とした。その後、これらのサムネイル画像から自動タグ付与に用いる被写体認識用の特徴ベクトルと類似画像検索のための特徴ベクトルがそれぞれ抽出される。このとき被写体認識については、各入力画像から抽出された特徴ベクトルに対して各タグモデルを適用し、ある閾値以上の画像についてはタグ名がインデックスとして付与される。なお、この時、被写体認識用の特徴ベクトルはどのタグモデルにも共通して使われるため、各画像から抽出された特徴ベクトルを保存しておくことで次の計算時間を節約できるように工夫した。

一方、類似画像検索用の特徴ベクトルについては、高速化のために、特徴を階層的に表現し、荒い分類からより詳細な類似性を比較する枠組みで検索できるようにした。

4.3 タグモデル管理システム

タグモデル管理システムは今回の提案において最も利用価値の高いものと考えている。これは、ユーザがシステム導入時には考えていなかった認識対象に対し、新たにそのタグモデルを作成し登録する機能を提供するものである。

図 9 は新たなタグモデルを作成するためのウェブアプリの画面例を示している。

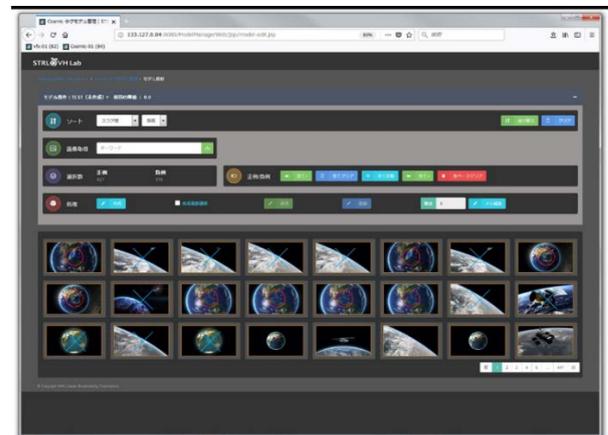


図 9 タグモデル作成のための画面例

Figure 9 Screen shot of application for creating an image recognition model.

このアプリ画面では下段に 3 列で合計 24 枚の画像が並んでいる。これは検索対象の画像全体のほんの一部であるが、ページを変えれば新たな 24 枚の画像が表示される。ここで、認識対象とする被写体（正例）に○印を、対象ではない被写体で積極的に排除したい画像（負例）には×印をつける作業を行う。この作業を、順次ページを切り替えながら進めていく。そして、ある程度の数に○印をつけたところで、そこまでの正例と負例を基にした学習によりタグ

モデルを作成する[5]. 今回は、タグモデル作成のための学習法として SVM (Support Vector Machine) を使った.

図 9 の上段には画像の並べ方 (認識スコア順やファイル順など) や、効率的に○×をつけるため (画像にすべて○印を付けるや、○印以外の画像すべてに×印をつけるなど) の指示を与える GUI となっている.

ここで、簡単な事例をもってタグモデルの作成手順をより詳しく説明する. 例としてマグロの握りずしを認識するタグモデルの作成手順を示す.

図 10 では、マグロの握りずしに○印がつけられ、それ以外には適当に×印がつけられている (この例では合計 25 枚の画像が表示されている). 本システムでは、設定した数になるだけ自動的に負例が抽出されるので、必ずしも×印を明示的につける必要はない. ×印を明示的につける理由は、対象とする画像に似ているが正例ではない画像に×印をつけた場合、より性能の良くなることがあるためである.

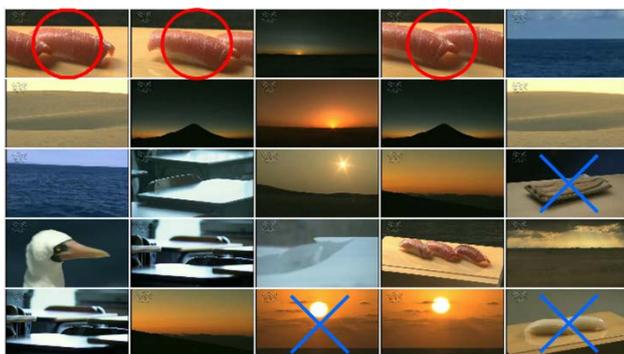


図 10 正例 (○印) と負例 (×印) の例

Figure 10 Example of selecting positive examples(○) and negative examples(×).

次に、選んだ正例と負例を用いて SVM によりマグロの握りずしのタグモデルを作成する. そしてタグモデル作成後に、このモデルをすべての検索対象画像に適用し、確からしさの値の高い順に並べ直して表示する. 図 11 に並べ直した画面例を示す (左上から右下に高い順に表示).

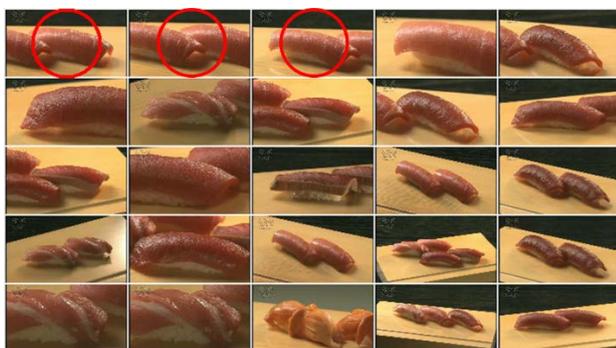


図 11 タグモデル適用後の画面

Figure 11 Result of applying image recognition model to all images in a database.

図 11 をみると、先ほど○印をつけた画像が上位にきてい

ることがわかる. また、○印はついていないがマグロの握りずしの画像が上位を占めていることもわかる. この結果では、最下段中央に異なるすしネタの画像が入っているので、例えば、これに再度×印をつけ、残りすべて (マグロの握りずし) に○印をつけてタグモデルを学習し直すと、さらに認識精度の高いタグモデルを作ることができる. このような作業を数回繰り返すことでより精度の高いタグモデルを獲得できる. 満足するところまで繰り返しの学習を進めたら、そのときのタグモデルを閾値も決めて登録する.

以上のように、システム導入後に簡単に新しいタグモデルを追加できることが、今回のシステムの大きな特長である. この追加機能は、MPF のアーキテクチャを採用していることで、容易に実現できた.

なお、最初のタグモデルを作成するときには、正例とする画像を見つけることが一般的に難しい (今回の写真検索でも対象は 13 万枚を越えていた). そこで、初回のタグモデル作成時には、ネットの画像検索サービスを利用して、自動的に正例候補の画像を収集できる機能も組み込んだ.

4.4 検索システム

インデックス作成後には、いくつかの方法で検索ができるようになる. 以下に提供している検索方法を列挙する.

- 自動で付与されたタグで検索.
- 画像をクエリとした類似画像検索.
- 事前に取り込んだ素材関連情報による検索.
- 検索システム利用時に入力したメモによる検索.

また、その他の機能として次のものも用意した.

- 一部は AND 操作が可能.
- 自動で付与されたタグの修正やメモの追加.
- 検索により発見した素材のオリジナルデータの取得.

なお、検索の GUI は今回の 2 つのシステムで少し異なるため、以下では別々に紹介する.

(1) 写真素材管理の場合

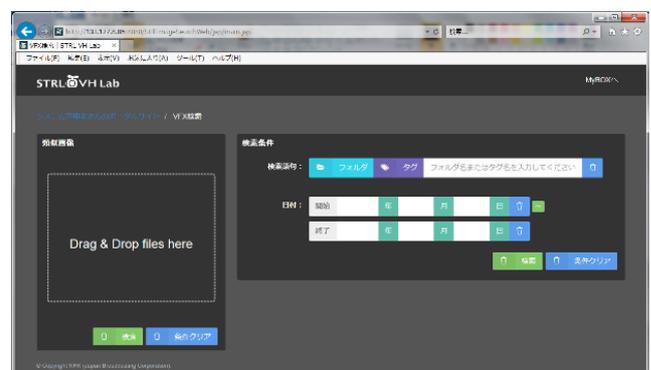


図 12 検索条件入力画面

Figure 12 Display for inputting retrieval conditions.

図 12 に写真素材の検索システムの検索条件の入力画面を示す. 左側の破線の四角い枠内は類似画像検索をしたい

場合にクエリー画像を選ぶための領域となっている。右側は、自動で付与されたタグ名、フォルダ名、メモから検索するための領域で、上段はそれらの検索語を入力する領域となっている。下段の2行にわたるところは年月日を入力する領域となっており、期間を区切って検索対象を絞れるようになっている。

図 13 には検索結果の表示例を示した。検索でヒットした画像が並べて表示される。

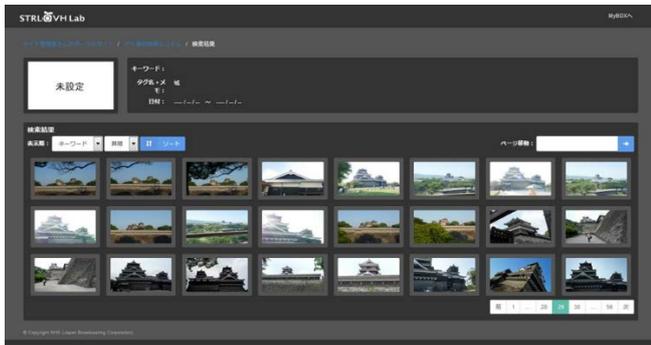


図 13 検索結果

Figure 13 Display of a retrieval result.

(2) 動画、静止画群など関連素材をグルーピングして管理する場合

宇宙関連番組のセクションでは、4K を含む実写映像、CG 映像、CG モデル、コマ撮り天体写真、音声素材、ドキュメントなど（これらを素材アイテムと総称する）、多岐にわたる素材が検索対象であった。但し、同じ科学事象などに対する、例えば CG 映像やコマ撮り写真、関連ドキュメントといった素材アイテムをグルーピングして検索したいという要望があったため、これらを一つのフォルダ（素材フォルダと呼ぶ）に入れ、これを検索の基本単位とした。

図 14 に、宇宙関連番組のセクションに向けて構築した検索システムの検索条件の入力画面を示す。



図 14 検索条件入力画面

Figure 14 Display for inputting retrieval conditions.

基本は VFX のセクションと同じであるが、この表示で

は自動付与したタグ名によって検索する GUI が、対応するボタン（この図では右上の細長い9個の四角）をクリックすることで検索できるようにしており、操作の簡略化を図ったバージョンとなっている。

また、手動で入力した素材関連情報から検索する場合、各項目をプルダウンメニューで選べるようにし、その項目に応じたテキストを入力できるようにした。



図 15 検索結果

Figure 15 Display of a retrieval result.

図 15 には、素材関連情報による検索結果の例を示す。ここでは、1 行毎に各素材フォルダの情報が表示されている（この表示ではトータル4行）。行の右側の縦長の楕円ボタンは、素材フォルダにある素材アイテムに該当するボタンの色が変わるようにしている。この例では、すべてが一番左の項目（CG/動画）のみが含まれる例になっている。

一方、類似画像検索の場合は検索結果が画像となるため、写真素材用のシステムと同様の表示方法とした（図 16）。

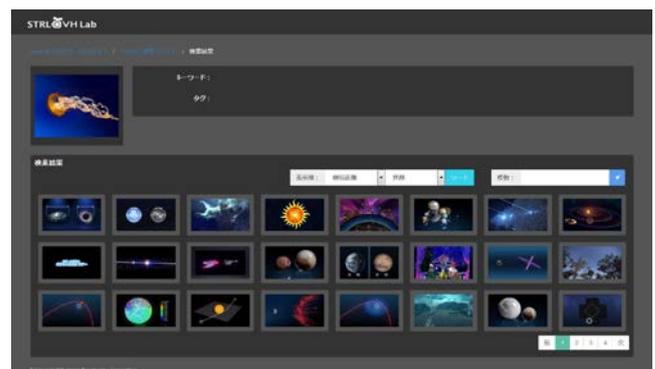


図 16 類似画像検索の結果

Figure 16 Display of result by similar image retrieval.

図 17 には素材フォルダの詳細を表示する画面を示した。ここでは、素材フォルダに入っている素材の詳細情報が提示される。画面の左側には番組情報が、右側には素材アイテム毎の情報が表示される。この例では素材アイテムが一つであるが、複数ある場合には右の欄に同様のパターンで

繰り返し表示されるようになっている。

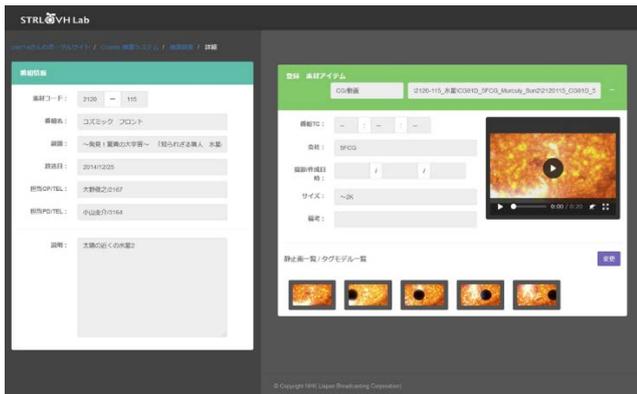


図 17 素材の詳細情報

Figure 17 Detailed information of materials.

図 17 の右下にある 5 枚のサムネイルは、素材映像の内容をすばやく把握できるようにするためのものである。今回対象とした宇宙関連の映像は、多くが例えば星空を撮影した動画など、時間変化が少ないため、最初と最後の 1 枚と、中間から等間隔に分割した 3 枚を抽出して、合わせて 5 枚を表示するようにした。なお、以前に開発した震災の素材映像に対するシステムでは、動画の扱いについては、まずカット替わりで動画をショットに分割し、これを素材検索の基本単位としていた。そして、分割したショットから代表サムネイルを、動画の時間変化が少ない部分からは少ない数で、変化が多い部分からは比較的多くを抽出するという技術[6]を使っていた。この技術は、速度の面でも、機能の面でも優れており、多数のシステムで利用されている。しかし、今回は、上記のような特殊事情があったため震災用に用いた技術は利用しなかった。

以上のように、検索のための画面構成等は、ユーザーの要望に応じて変更しているが、中核となるメタデータ生成までを共通化しているため、ユーザー毎にシステムを一から作り直す必要はない。今回のシステムで言えば、素材関連情報の取り込み方法と検索 GUI において各ユーザーへの適用化を図ればよかった。

5. おわりに

本稿では、放送現場で試用を開始した放送番組の素材検索システムについて主にその機能面から実例に即して紹介した。映像解析処理からメタデータ生成までを共通化し、タグ名を付与するためのタグモデルもシステム導入後にユーザーが容易に追加できるところに本システムの特長がある。

今後は、今まで想定していなかった現場からの要望も出てくるものと考えており、現在の試用において上げられてくる要望も合わせて考慮し、機能追加や性能改善を図ると共に、効率的に新規システムの構築が行えるような仕組みに改良を加えていきたいと考えている。

参考文献

- [1] 住吉英樹, 河合吉彦, 望月貴裕, 佐野雅規, 藤井真人. 大震災アーカイブス メタデータ補完システム. 映像情報メディア学会年次大会, 2012, 6-1.
- [2] 河合吉彦, 藤井真人. 映像からのオブジェクト識別技術. NHK 技研 R&D, 2013, No.142, p.43-51.
- [3] 望月貴裕, 住吉英樹, 藤井真人. クエリー画像描画による画像検索の構図テンプレートをを用いた高速化. FIT. 2010, No.3, H-040, p.219-220
- [4] 佐野雅規, 住吉英樹, 藤井真人, 柴田正啓, 八木伸行. メタデータ制作フレームワーク. NHK 技研 R&D, 2010, No.121, p.26-33.
- [5] Kawai, Y. and Fujii, M.. Semantic Concept Detection Based on Spatial Pyramid Matching and Semi-supervised Learning. ITE Trans. Media Technology and Applications. 2012, vol.1, no.2, p. 190-198.
- [6] 河合吉彦, 住吉英樹, 藤井真人, 八木伸行. 逐次的な特徴算出によるディゾルブ, フェードを含むショット境界の高速検索手法. 電子情報通信学会論文誌 D, 2008, Vol.66, No.11, p.444-452