検索質問と字幕の文字画像特徴量間の距離に基づく字幕検索手法

西川伸紀[†] 獅々堀正幹[††]
柘植覚[††] 北研二[†††]

本研究では映像内の文字情報である字幕に着目した字幕検索システムを開発する．従来，字幕検索は映像内に出現する字幕に対して文字認識を行う手法が主流であった．しかし，この手法では，事前に文字認識を行うための時間コストが必要であり，また，完全な文字認識結果が得られない場合には検索精度が低下するという問題があった．本論文では，上記の問題点を解決した高精度かつ高速な字幕検索手法を提案する．字幕検索を実現するためには，映像中に出現するすべての字幕を正確に認識する必要はなく，検索キーに対する字幕だけを認識できれば適切な検索結果を得ることができる．そこで本手法では，各字幕の文字画像特徴量と検索キーに対応する文字画像特徴量との距離に基づいて該当の字幕が出現するフレームを検索する．また，各字幕の文字画像特徴量を多次元索引化することで，検索キーの文字画像特徴量との距離計算を高速化する．さらに，本手法では検索過程で特徴量照合を行うため，前処理で文字認識処理が必要でなく，時間コストを軽減することができる．実際に 3 時間分の映像データに対して映像中の出現頻度が比較的多い 91 単語を用いて検索実験を行った結果，1-gram 特徴量を用いた場合には最大 98.61%，2-gram 特徴量を用いた場合には最大 99.59% の平均適合率を得ることができた．検索時間に関しても，2-gram 特徴量を用いた場合でも約 0.5 秒で検索結果を得ることができた．

A Method to Retrieve Video Telop Based on the Distance of Character Image Features between Query and Telop

Nobuki Nishikawa,[†] Masami Shishibori,[††] Satoru Tsuge[††]
and Kenji Kita[†††]

Video telop retrieval methods based on telop characters can retrieve the corresponding telops to the query from the huge video data. The conventional methods make the text data from the image data of telop characters by recognizing all telop characters in the video data, and then the full text search is operated toward the recognized text data. The conventional methods can not retrieve with high precision, because all telop characters can not recognize as their right characters perfectly. In this paper, a new video telop retrieval method based on telop characters is proposed. In order to specify the suitable telop, this method recognizes the only corresponding telop characters to the query keyword not all characters. This method calculates the distance between each image features of telop characters and template image features of query keyword. The number of distance calculations can decrease by indexing the multidimensional data for image features of telop characters. Experimental results, using 91 query keywords, show that the average precision of proposed method using 1-gram feature becomes 98.61%, and using 2-gram feature becomes 99.59%. Moreover, the retrieval time can be obtained in about 0.5 seconds when using 2-gram feature.

† 徳島大学大学院工学研究科
Graduate School of Advanced Technology and Science, the University of Tokushima
†† 徳島大学大学院ソシオテクノサイエンス研究部
Institute of Technology and Science, the University of Tokushima
††† 徳島大学高度情報化基板センター
Center for Advanced Information Technology, the University of Tokushima

1. はじめに

近年ネット配信システムの発展にともなって大量の映像が我々の身の回りに溢れている．また，大容量の記憶メディアが普及し，ユーザが映像を保存し，映像ライブラリを構築する環境が整いつつある．そのため，保存された大量の映像から必要な映像シーンを検索するために内容に基づいた映像検索技術（Content-

Based Video Retrieval）が近年おおいに注目されている[1),2)]．映像には時間経過にともない内容が変化するという特徴がある．よって，映像の一部分を見ただけで目的のシーンが含まれている映像であるかを判断することは難しく，映像全体を確認する必要があるため，大量の映像から必要な映像を選択するためには膨大な労力を要する．そこで，自動的に映像内容を把握して必要な映像シーンを取得できれば労力を軽減することが可能である．

映像は音声，画像，文字等の複数の情報から構成されているが，本研究では映像内の文字情報である字幕に着目する．字幕はその表示方法によりクローズドキャプションとオープンキャプションに分けられる．クローズドキャプションは，字幕放送のことであり表示には専用のデコーダが必要である．また，字幕内容はテキストデータで受信され映像に重ねて表示される．一方，オープンキャプションは画面上に常時表示されている字幕で，特別な機器がなくても表示可能であり，ニューステロップや，映画，ドラマの字幕等はオープンキャプションである．一般に，家庭で録画される字幕つき映像はオープンキャプションであるためテキストデータが付与されておらず字幕を文字認識しなければ検索システムに応用することができない．そのため，本研究ではオープンキャプションを対象とする．特に，オープンキャプションの中でもドラマやアニメーション等の登場人物の音声情報を字幕として表示したものを対象とし，映像中の字幕文字を解析することで検索要求を含む字幕を検索するシステムを開発する．これらの字幕は，映像の意味的内容と関連している場合が多く，また，シーンの開始と同時に表示されるため映像と時間的にも密接な関係がある場合が多いことから，検索要求に適合したシーンを取得するために非常に有効である[3)〜5)]とされ，字幕検索はシーン検索への応用が期待できる．

従来の字幕検索手法は，字幕の書きおこしテキストを人手で用意し，これに対して全文検索を行う手法と字幕画像の文字認識結果を利用する手法に大別される．書きおこしテキストを用いる手法は，高い検索精度を得ることができるが，作業者の負担が大きく大量の映像に適用するには不向きであった．一方，文字認識を用いる手法は，字幕画像に文字認識を適用することで字幕画像のテキスト変換が自動化されるため，大量の映像にも適用可能である．しかし，事前に文字認識を行うための時間コストが必要となり，また，完全な文字認識結果を得ることが難しいため正確なテキストが作成されず検索精度が低下することが問題であった．

本論文では，上記の問題点を解決した高精度かつ高速な字幕検索手法を提案する．まず高精度化への改良として，字幕の文字を認識することと字幕検索の違いに着目する．字幕認識では，すべての字幕内の文字画像を認識する必要があるが，字幕検索では検索キーワードが出現する字幕のみが特定できればよいため，検索キーワードの出現している部分だけを認識できれば適切な検索結果を得ることができる．そこで本手法では，検索キーワード内の文字が映像内のどの字幕内に出現しているかを字幕内の文字画像特徴量と検索キーワードに対応する文字画像特徴量との距離計算に基づいて決定する．

また，高速化を実現するために，多次元ベクトルで表現された文字画像特徴量を多次元索引化[6)]することで，検索キーワードの文字画像特徴量との距離を近傍検索し，検索キーワードが出現している字幕を高速に特定する．さらに，検索精度を向上させるため，隣接する文字画像間には特徴の関連性があると考えられることから，隣接する 2 文字の文字画像特徴量からなる 2-gram 特徴量を導入し精度向上を図る．

以下，2 章では従来の字幕検索手法を紹介し，それらの問題点を明確にする．3 章では本研究の着目点を述べた後，本提案手法の概要について説明する．4 章では本手法の各モジュールの詳細を示し，5 章では本手法を 2-gram 特徴量に拡張する方法を述べる．6 章において，本手法の有効性を検証するため，評価実験を示し，その考察を述べる．最後に 7 章において，まとめおよび今後の研究課題について述べる．

2. 従来の字幕検索手法

従来の字幕検索手法には，書きおこしテキストを用いる手法と文字認識を用いる手法が主に使用されてきた．書きおこしテキストを用いる手法は，人手で字幕画像をテキストに変換できるため正確な書きおこしテキストが作成でき，高い検索精度を得ることができるが，作業者の負担が大きいという問題点がある．書きおこしテキストを作成する労力の負担を軽減する手法として文字認識を用いる手法がある．文字認識を用いた字幕検索手法の概要を**図 1** に示す．また，以下に文字認識を用いた字幕検索手法の手順を示す．

- **手順 1： 字幕領域の切り出し**[3),7),8)]
  映像から字幕の切り替わり点を検出し，字幕領域を特定し字幕部分を切り出す．さらに，切り替わり点の時間をタイムスタンプに記述する．
- **手順 2： 文字画像の切り出し**[9),10)]
  字幕画像を 2 値化して文字列画像を切り出し，文

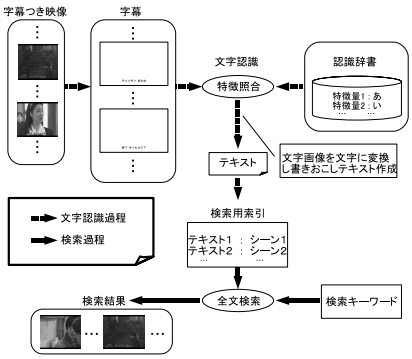

図 1 文字認識を用いた字幕検索手法の概要
Fig. 1 Outline of the conventional method.

字列画像から文字画像を切り出す．

- **手順 3： 文字画像特徴量による文字認識**[11)〜13)]
  切り出した文字画像から文字画像特徴量（以下，単に特徴量と記す）を抽出し，認識辞書と特徴量照合し文字画像の文字を特定する．
- **手順 4： 字幕のテキスト化**
  特定された文字をファイルに書き出し，書きおこしテキストを作成する．
- **手順 5： 索引の作成**
  書きおこしテキストの文字とその文字の出現する位置を対応させて検索用索引を作成する．

［手順終了］

検索の際には，全文検索を行い検索キーワードが含まれている字幕位置を取得することで，字幕を決定しタイムスタンプの時間情報をもとに対応した字幕が表示されているフレームを表示する．この手法では字幕文字が正確に認識できれば全文検索によって正確な字幕検索が可能であるが，文字認識の結果に誤認識が含まれていた場合は，誤った検索用索引が作成されるため正確な字幕検索ができなくなる．

3. 文字画像特徴量間の距離に基づく字幕検索手法

3.1 着 目 点

2 章で述べた手法では，字幕内の文字画像を文字認識する際に誤認識が起こり全文検索での検索精度低下の原因となっていた．文字認識では，抽出した文字の特徴量と文字テンプレート（認識辞書）を照合し文字画像がどの文字であるかを識別する．一般に文字テンプレートには，数万単位の文字がその特徴量と対応づけて格納されている．この文字テンプレートに特徴量の類似した文字が多く含まれていると，誤認識が発生する確率が高くなる．

表 1 誤認識した文字種の例
Table 1 Example of character class which fail to recognize.

正解文字	句	初	間	手	ウ	時	日	ヒ
誤認識文字	句	祝	問	チ	ワ	暗	目	と

ここで，類似した文字種の文字認識に対する影響を確認するため字幕に対する文字認識実験を行った．使用した映像データは 60 分ドラマ 3 話分を用いた．また，文字テンプレートとして 60 分ドラマ 9 話分の字幕の文字画像を用いて人手で作成した文字テンプレート（文字種数は 1,071 種類）を使用した．実験の結果，認識精度は 97.85% であった．また，誤認識を発生していた文字は 1,071 種類中 60 種類であった．**表 1** に誤認識の例を示す．このことから，文字テンプレート内の文字種が 1,000 種類程度でも類似文字の影響が現れることが分かる．大規模な文字テンプレートを用いると認識精度がより悪化することが予想される．特に誤認識した文字については，後で全文検索をした際に検索もれを生じるため再現率が低下してしまう．たとえば，「目」を文字候補 3 件で文字認識する場合，「自，日，白」と誤認識すると検索では「目」を検索することはできない．このとき，文字テンプレート内に類似した文字が多いほど候補文字の中に正解文字が含まれない確率は高くなる．

字幕検索で重要な点は，検索キーワードを含む字幕をもれなく検出することであり，そのためには，再現率を向上させる必要がある．また，字幕検索を行うにはすべての字幕を認識する必要はなく，検索キーワードがどの字幕に出現しているかが分かりさえすれば字幕検索ができる点に着目する．再現率を向上させるために，検索キーワードを構成している文字を認識することだけを目的とし文字テンプレート内の類似文字の誤認識を防ぐ，すなわち，検索キーワードを構成している文字の文字テンプレートだけを使用して文字認識を行えば再現率は向上すると考えた．

しかし，検索キーワードのテンプレートだけを用いた場合，今度は字幕中の類似文字が検索キーワードを構成している文字と誤認識され，字幕が過剰に検索されるため適合率が低下する．そこで，適合率を向上させるため，検索キーワードの文字の連続性を利用する．検索キーワードを構成する文字は同一字幕内で連続かつ順番どおりに出現することで検索キーワードと同じ文字列を構成する．そのため，検索結果に文字列の連続性がないものを削除することにより，適合率を向上させることができる．その結果，再現率が 100% に近く，適合率も高い字幕検索ができる．

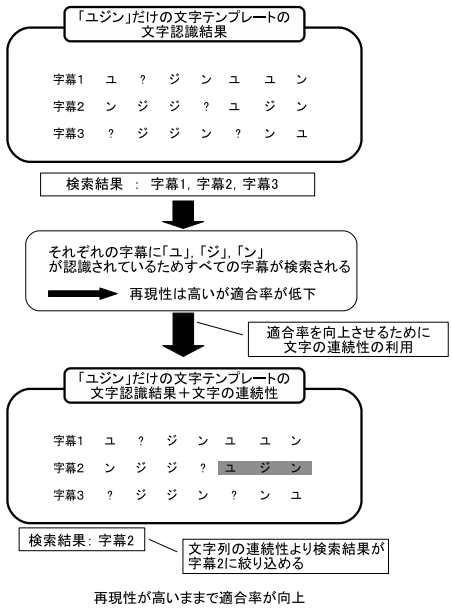

図 2 着目点
Fig. 2 Outline of an idea of the proposed method.

以上の考え方を**図 2** に示し，以下で図 2 を用いて説明する．まず，検索キーワード「ユジン」が入力された場合,「ユ」,「ジ」,「ン」の文字テンプレートにだけ着目する．このテンプレートを用いて字幕 1，字幕 2，字幕 3 に対して文字認識を行うと，認識結果は図 2 上段のように認識される．ここで，図中の？は認識されなかった文字を表す．字幕 1〜字幕 3 それぞれに「ユ」,「ジ」,「ン」の文字が認識されているため，検索結果として，字幕 1〜字幕 3 を出力する．その結果，再現性は高くなるが，字幕 1，字幕 3 には文字列「ユジン」が含まれていないため，過剰に検出され，適合率は低下してしまう．そこで，適合率を向上させるため，検索キーワード「ユジン」の文字の連続性を利用する．「ユ」,「ジ」,「ン」が同一字幕内にかつ，連続して出現している字幕を検索することで，検索結果を字幕 2 に絞り込むことができるため，再現性が高いままで適合率が向上する．

3.2 概　　要

3.1 節で述べた考えに基づく手法の概要を**図 3** に示し，本手法の処理手順を以下に述べる．まず，検索対象となる映像データに対して行われる前処理の手順について述べる．

- **前処理手順 1： 従来法の手順 1 と同様**
- **前処理手順 2： 従来法の手順 2 と同様**
- **前処理手順 3： 特徴量抽出**

  切り出した文字画像から特徴量を抽出する．

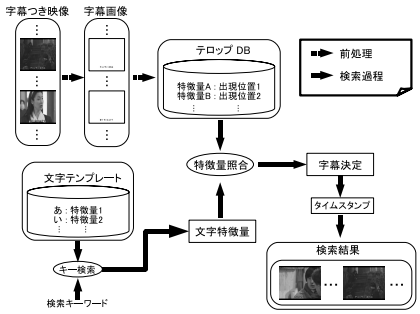

図 3 文字画像特徴量間の距離に基づく字幕検索手法の概要図
Fig. 3 Outline of the proposed method.

- **前処理手順 4： テロップ DB の作成**

  抽出した特徴量とその出現位置を対応づけてデータベースを作成する．以下このデータベースをテロップ DB と呼ぶ．

［前処理手順　終了］

次に検索の手順を述べる．

- **検索手順 1： 検索キーワードの分割**

  入力された検索キーワードを 1 文字ごとに分割する．

- **検索手順 2： キー検索による特徴量の取得**

  分割された文字を検索キーとして文字テンプレートからその文字に対応する特徴量を取得する．

- **検索手順 3： 特徴量照合**

  文字テンプレートから取得した特徴量と文字画像から抽出した特徴量との特徴量照合を行う．

- **検索手順 4： 字幕の決定**

  検索キーワード内の各文字の特徴量照合の結果を類似度の高い順（距離の小さい順）に組み合わせることにより，検索キーワードが含まれている字幕を決定する．字幕の決定手法については 4.4 節で詳しく述べる．

［検索手順　終了］

本手法を用いることで，特徴量の照合回数は,（映像内に出現する特徴量の数）×（検索キーワードの文字数）回となり，照合コストを低く抑えることが可能になる．一般にドラマ 1 話分，約 60 分の映像中にはおよそ 500 種類の文字が含まれている．検索キーワードの文字数はたかだか 10 文字程度であると考えられるため，提案手法を用いることでドラマ 1 話分の映像を検索するためには，およそ 500×10 回の照合回数でよいことになる．

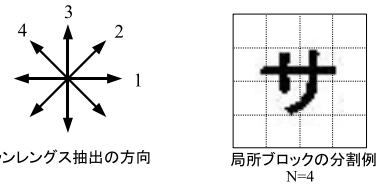

図 4 ランレングスの方向と領域分割例
Fig. 4 Example of division and scanning direction.

4. 各モジュールの詳細

4.1 文字画像特徴量

本手法では，文字画像特徴量として方向寄与度特徴[13] を用いた．方向寄与度特徴とは文字の方向情報である文字線のランレングスを特徴として用いたものである．文字線のランレングスとは，ある画素に対して文字を構成している画素の**図 4** に示す横方向，右斜め方向，縦方向，左斜め方向の 4 方向に対する連続性を示すものである．方向寄与度特徴の算出手順について説明する．

・手順 1： 局所ブロックへの分割
入力画像を $N \times N$ のブロックに分割する．

・手順 2： 画素に対するランレングスの計算
分割された局所ブロック内の各黒画素に対してランレングス l_i（$i = 1, 2, 3, 4$）を求める．

・手順 3： 局所ブロックに対するランレングスの計算
m（$1, 2, \cdots, N \times N$）ブロックごとにランレングスを平均し，局所ブロックに対するランレングス $l_{m,i}$ を求める．

ここで，l_1，l_2，l_3，l_4 はそれぞれ各黒画素における横方向，右斜め方向，縦方向，および左斜め方向のランレングスを表す．また，入力画像を分割して得られる局所ブロックを m（$1, 2, \cdots, N \times N$）とし，第 m ブロック内で算出される方向寄与度を $d_{m,i}$ とし，$l_{m,i}$ を第 m ブロックにおいて l_i を平均化して得られるランレングスとする．**図 5** に入力画像の分割例，およびランレングスの抽出法を示す．また，方向寄与度特徴の計算式を式 (1) に示す．

$$d_i = \frac{l_i}{\sqrt{\sum_{j=1}^{4} l_j^2}} \tag{1}$$

4.2 文字テンプレート

文字テンプレートはテキストとして入力された文字からキー検索技術により特徴量を取得するための特徴量データベースである．文字テンプレートは文字とその特徴量を対応づけて登録する．

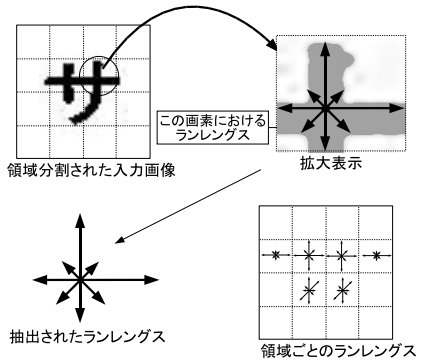

図 5 方向寄与度特徴
Fig. 5 Example of DC extraction.

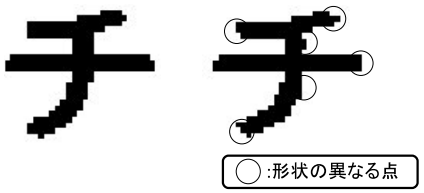

図 6 同じ文字の文字画像の形状のずれの例
Fig. 6 Difference of shape of the same characer.

図 6，および 3.1 節で述べたように字幕中の文字画像は同じ文字であっても出現位置が異なれば文字の形状が異なっている場合がある．そのため，文字テンプレート作成の際には，同じ文字の形状の相違を考慮し，特徴量のずれを可能な限り少なくする必要がある．そこで，文字テンプレートは映像内に出現するすべての同じ文字の特徴量を平均して作成する．なお，作成したテンプレートはキー検索を可能にするためにパトリシアトライ法[14] を用いてデータベース化しておく．

4.3 テロップ DB

テロップ DB は映像中の字幕文字から各文字画像の特徴量を抽出し，その特徴量と各文字の出現位置とを対応づけたデータベースである．たとえば「014-03」（14 番目の字幕の 4 番目の文字）の文字画像とその特徴量「特徴量 A」は「014-03:特徴量 A」のように登録する．テロップ DB 作成の流れを**図 7** に示す．また，特徴量の照合には高速最近傍検索アルゴリズム[15] を用いるため，作成したデータベース内の特徴量に対してあらかじめデータ変換を行っておく．

4.4 検索結果の絞り込み

図 8 に検索キーワードの出現位置および類似度の決定手法の概要を示す．検索の際には，検索キーワードの各文字の特徴量とテロップ DB 内の特徴量とを照合し，検索キーワードの各文字が出現する字幕位置を特定する．ここで，検索キーワードの各文字は連続し

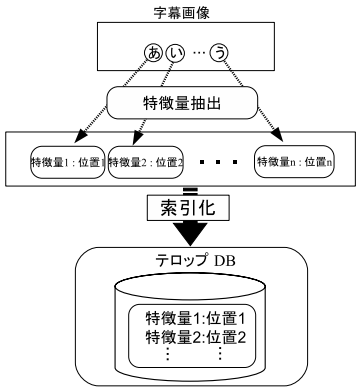

図 7 テロップ DB 作成の流れ
Fig. 7 Outline of making a telopDB.

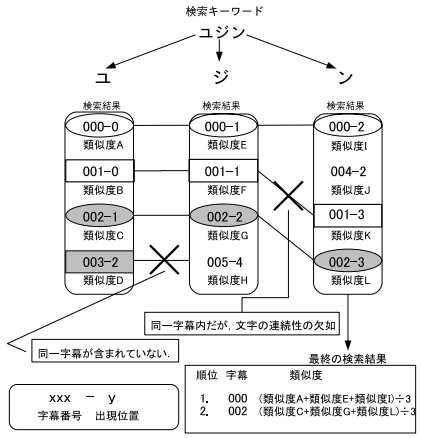

図 8 絞り込み手法および類似度の算出法
Fig. 8 Calculation of the similarity.

て出現するため，分割された検索キーワード内の文字が同一の字幕内に存在し，かつ，検索キーワードと同じ順番で連続して存在することを（検索キーワードに相当する）字幕が適切に存在するための条件（字幕特定条件）とする．つまり，1 文字目の検索結果（字幕位置）から n 文字目までの検索結果（字幕位置）の共通部分，かつ，連続性を考慮することにより，検索キーワードを含む字幕位置を決定する．

また，本手法は特徴量どうしの距離計算を行っているため，各文字画像特徴量間の距離が算出される．この距離が小さいほど文字の類似度が高いと見なすことができる．この距離を各文字画像特徴量間の類似度として用いる．そして，特徴量間の距離の小さい順，すなわち，類似度の高い順に検索結果をソートすることで，検索結果の上位に類似文字が現れる．そして，1 文字ごとに算出される類似度のうち，同一フレームに存在する字幕に対する類似度を加算平均することで字幕内の文字列に対する検索キーワードの類似度を算出し検索キーワードを含む字幕の順位づけを行う．

検索結果の絞り込み手法と類似度の算出法について図 8 を用いて詳しく説明する．検索キーワードとして「ユジン」が入力されたとき，まず，最初にキーワードを「ユ」,「ジ」,「ン」のように各文字に分割する．そして，各文字の検索を行い検索結果を取得する．この検索結果を字幕特定条件に従って組み合わせる．「ユ」の検索結果の「000-0」,「ジ」の検索結果の「000-1」,「ン」の検索結果の「000-2」は同一フレームに存在する字幕上に存在し，連続した文字であるため最終の検索結果として出力する．ここで，検索結果の「xxx-y」の xxx は字幕番号であり，y は文字の字幕内での出現位置である．つまり,「000-0」とは「000」番の字幕の一番最初の文字である．このとき，類似度は「類似度 A」,「類似度 E」,「類似度 I」を可算平均した値とする．同様に「001」番の字幕を見てみると「ユ」,「ジ」には連続した文字が検索されているが,「ン」では文字の連続性が欠如しているので最終の検索結果として出力しない．また,「002」番の字幕はすべての検索結果に含まれ，文字の連続性が確保されているため最終の検索結果として出力する．このときの類似度は 000 番の字幕の場合と同様に「類似度 C」,「類似度 G」,「類似度 L」の可算平均であるが，各文字の検索結果は類似度の高い順にソートされているため「000」の類似度 >「002」の類似度である．また,「003」,「004」,「005」番の字幕については同一フレームに存在する字幕が検索されていないため，最終の検索結果として出力しない．最後に最終の検索結果（検索キーワードが含まれる字幕の候補）として，1 位に「000」，2 位に「002」を出力する．

5. 2-gram 特徴量を用いた字幕検索手法

3.2 節で述べた手法では 1 文字単位の特徴量を用いているため映像中に検索キーワードと類似した文字が多く存在している場合，検索精度が低下するという問題点があった．さらに，1 文字単位の出現位置の連続性を用いて字幕を決定しており，隣接文字の特徴の関連性が考慮されていなかった．検索精度をさらに向上させるためには，1 文字単位の出現位置の連続性だけでなく，隣接する 2 文字単位の特徴を考慮する必要がある．そこで，文字画像特徴量を 2-gram 文字の考え方を用いて拡張する．2-gram 文字特徴量は，隣接する 2 文字の特徴量をまとめて 1 つの特徴量を作成することで，特徴量に隣接文字間のつながり情報と出現する順序情報を持たせる．以下，3.2 節で用いた 1 文字

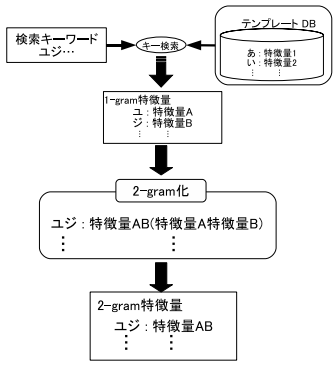

図 9 2-gram 化の手順
Fig. 9 Outline of generating 2-gram feature.

分の特徴量を 1-gram 特徴量と呼び，隣接する 2 文字分の特徴量を 1 つの特徴量にまとめたものを 2-gram 特徴量と呼ぶ．2-gram 特徴量を用いた手法の処理手順を以下に示す．まず，前処理の手順を示す．

- **前処理手順 1〜3： 1-gram での前処理手順 1〜3 と同様**
- **前処理手順 4： 2-gram 特徴量の作成**
  1 文字目の特徴量の後に後接する文字の特徴量を付加することにより 2-gram 特徴量を作成する．
- **前処理手順 5： 1-gram での前処理手順 4 と同様**

［前処理手順 終了］

次に検索の手順を示す．

- **検索手順 1〜2： 1-gram での検索手順 1〜2 と同様**
- **検索手順 3： 特徴量の 2-gram 化**
  取得した 1 文字目の特徴量の後に後接する 1 文字の特徴量を付加することにより 2-gram 特徴量を作成する．2-gram 化の手順を**図 9** に示す．
- **検索手順 4： 特徴量照合**
  検索キーワードの 2-gram 特徴量とテロップ DB 内の特徴量との特徴量照合を行う．
- **検索手順 5： 1-gram での検索手順 4 と同様**

［検索手順 終了］

前処理手順 4，検索手順 3 の過程を付け加えることにより 1-gram 特徴量を用いたシステムに大きな変更を加えることなく 2-gram 特徴量を用いたシステムに拡張することが可能である．

6. 評 価

6.1 実験方法

検索に使用する映像データとして，字幕の使用文字が未知の 60 分ドラマ 3 話分を使用した．ただし，今回の実験では，背景画像が文字画像に与える影響を可能な限り無視するため，映像データが格納された DVD のサブピクチャ（背景が透過処理された字幕画像）から抽出した字幕を使用し文字テンプレートを作成した．また，字幕から文字画像の切り出しが完全に行われたと仮定して実験を行った．これらの処理を行ったのは，今回の実験の目的が本手法の改良による検索精度の向上を純粋に検証するためである．

まず，1-gram 特徴量として方向寄与度 256（4 方向×64 分割）次元，および 2-gram 特徴量として方向寄与度 512（256+256）次元を使用して特徴量を作成した．また，文字テンプレート作成に使用した映像データは，60 分ドラマ 9 話分の映像を用い映像中に出現した同じ文字の特徴量を平均して作成し，1 話から 9 話まで 1 話分ずつ増加させた 9 種類のテンプレートを作成した．同じ文字の特徴量を平均したのは同じ文字の文字画像の形状が異なっているためであり，文字種数を変化させたテンプレートを作成したのは，文字種の変化が及ぼす影響を検証するためである．また，文字テンプレートは文字とその特徴量に対して正しい対応づけが必要であるため，書きおこしテキストを用いて対応づけを人手で確認して作成した．

検索キーワードは，実際の映像データ中に出現する単語のうち出現頻度が 4 回以上の 91 語を使用した．また，同一の文字テンプレートを用いて 3.1 節で用いた文字認識を適用して，書きおこしテキストを作成し，同様の検索キーワードを用いて検索を行い提案手法の精度と比較した．

評価尺度としては，検索結果の再現率，平均適合率，F 値[14] を用いて評価した．また，映像中の字幕における検索語句の出現の有無を正解データとして用いた．

また，同一の検索キーワード（91 語）を用いて各手法の前処理の時間，および検索時間を比較した．さらに，提案手法の映像データ量に対する時間評価を行うために 60 分ドラマ 3 話分と 60 分ドラマ 30 話分の映像データを使用し，各処理部の所要時間を計測した．ここで，各処理部とは 3.2 節および 5 章での検索手順の「キーワード分割」,「キー検索による特徴量取得，作成」,「特徴量照合」,「字幕決定」の各手順のことである．なお，今回の実験では文字画像の切り出しは，完全に行われたと仮定して評価しているため，文字画像の切り出し時間は評価には加えていない．また，時間の測定に用いたマシンのスペックを以下に示す．

- CPU Intel(R) Pentium4 3.20 GHz
- メモリ PC-4200 DDR2 SDRAM 1 G byte

6.2 実験結果

まず，映像の話数を増加させたときのテンプレート

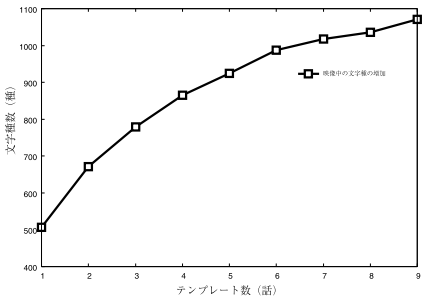

図 10 話数を増加させた場合のテンプレート内の文字種数の変化
Fig. 10 The number of character kinds in each movie.

表 2 各テンプレートに対して 1-gram 特徴量を用いた検索精度
Table 2 Experimantal results of the proposed method using each templates of 1-gram feature.

テンプレート	再現率 (%)	平均適合率 (%)
1-3	98.83	98.50
1-4	98.78	98.45
1-5	98.92	97.61
1-6	98.92	98.61
1-7	98.71	97.38
1-8	98.62	98.30
1-9	98.92	98.61

表 3 各テンプレートに対して 2-gram 特徴量を用いた検索精度
Table 3 Experimantal results of the proposed method using each templates of 2-gram feature.

テンプレート	再現率 (%)	平均適合率 (%)
1-3	100.00	99.59
1-4	100.00	99.57
1-5	100.00	99.58
1-6	100.00	99.59
1-7	100.00	99.59
1-8	100.00	99.59
1-9	100.00	99.57

内の文字種数の変化を**図 10** に示す．1-gram 特徴量を用いた各テンプレートに対する検索精度を**表 2** に示す．2-gram 特徴量を用いた各テンプレートに対する検索精度を**表 3** に示す．ここで，表内のテンプレートの値は，文字テンプレート作成に用いた最初と最後の話数を示す．たとえば,「1-5」の値は，1 話から 5 話までの映像内の文字画像を用いて作成したテンプレートを意味する．また，再現率は検索で用いた 91 単語の再現率の平均であり，平均適合率は各検索キーワードにおける平均適合率を平均したものである．

比較手法として，文字認識の結果から作成した書きおこしテキストを用いた検索精度を評価した．なお，各文字に対して文字認識を適用し，認識結果の上位 n 件を文字侯補として書きおこしテキストを作成した．

たとえば「ユジン」という字幕があり，$n = 2$ のときの文字認識結果が {ユ，ュ}，{ジ，ン}，{ン，y} であったとき,「ユジン」「ユジ y」「ユンン」「ユン y」「ュジン」「ュジ y」「ュンン」「ュン y」の 8 通りのテキストがあるものと見なす．$n = 1$，2，3 と変化させた際の検索精度をそれぞれ**表 4**，**表 5**，**表 6** に示す．

表 4 文字認識結果上位 1 位の文字に対する検索精度.
Table 4 Experimental results of the conventional method (1 candidate character).

テンプレート	再現率の平均 (%)	適合率の平均 (%)
1-3	89.98	100.00
1-4	91.29	100.00
1-5	91.60	100.00
1-6	91.87	100.00
1-7	91.50	100.00
1-8	93.81	100.00
1-9	94.33	100.00

表 5 文字認識結果上位 2 位までの文字に対する検索精度
Table 5 Experimental results of the conventional method (2 candidate characters).

テンプレート	再現率の平均 (%)	適合率の平均 (%)
1-3	99.06	100.00
1-4	99.06	100.00
1-5	98.74	100.00
1-6	98.74	100.00
1-7	98.74	100.00
1-8	98.74	100.00
1-9	98.74	100.00

表 6 文字認識結果上位 3 位までの文字に対する検索精度
Table 6 Experimental results of the conventional method (3 candidate characters).

テンプレート	再現率の平均 (%)	適合率の平均 (%)
1-3	99.41	98.81
1-4	99.41	98.93
1-5	99.27	99.02
1-6	99.27	99.02
1-7	99.41	99.02
1-8	99.38	98.98
1-9	99.22	98.79

用いるテンプレートを 1-9 に固定し，従来手法における文字認識候補を 1〜7 件に変化させ，提案手法（2-gram 特徴量）と精度比較した再現率・適合率曲線を**図 11** に示す．なお，図中の数字は文字認識候補数を表す．各手法の F 値を**表 7** に示す．ただし，文字認識手法の F 値は最も精度の良かった文字候補上位 2 件の結果である．また，時間評価として，各処理の所要時間手法の前処理および検索の所要時間を**表 8** に示す．最後に，提案手法の検索過程の処理時間の詳細を**表 9**，**表 10** に示す．ただし，使用する文字テンプレートは 1-9 を使用した．

表 9 提案手法における処理時間の詳細（3 時間）
Table 9 The details of processing time of the proposed method (3 hours).

	総時間 (s)	分割 (s)	特徴量作成 (s)	照合 (s)	字幕決定 (s)
1-gram	0.269	0.006	0.088	0.010	0.143
2-gram	0.525	0.006	0.194	0.004	0.291

表 10 提案手法における処理時間の詳細（30 時間）
Table 10 The details of processing time of the proposed method (30 hours).

	総時間 (s)	分割 (s)	特徴量作成 (s)	照合 (s)	字幕決定 (s)
1-gram	1.111	0.007	0.094	0.010	0.971
2-gram	2.693	0.006	0.193	0.003	2.454

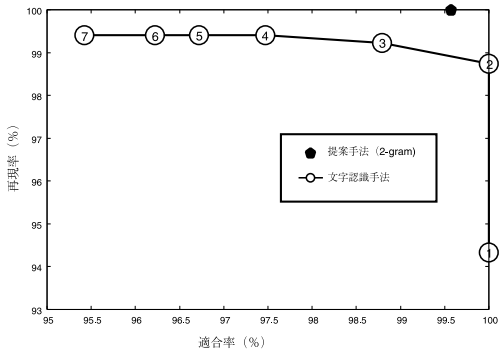

図 11 従来手法と提案手法との精度比較
Fig. 11 Retrieval precision compared the conventional method with the proposed method.

表 7 各手法に対する F 値
Table 7 F-measure of each method.

テンプレート	1-gram	2-gram	文字認識
1-3	0.9866	0.9979	0.9953
1-4	0.9861	0.9978	0.9953
1-5	0.9826	0.9979	0.9937
1-6	0.9876	0.9979	0.9937
1-7	0.9804	0.9979	0.9937
1-8	0.9846	0.9976	0.9937
1-9	0.9876	0.9978	0.9937

表 8 各手法の前処理および検索時間
Table 8 Cpu-time of preprocessing and retrieval of each method.

	前処理 (s)	文字認識処理 (s)	検索時間 (s)
文字認識	409.47	464.00	0.020
1-gram	409.47	—	0.269
2-gram	409.47	—	0.525

6.3 考　　察

まず，文字テンプレート内の文字種数を増加させたときの検索精度について考察する．表 2 から表 6 より，テンプレート内の文字種数を増加させても各テンプレートにおける検索精度にほとんど差は現れなかった．提案手法では，検索キーワード内の各文字の特徴量をパトリシアトライを用いて（文字をキーとした）キー検索により取得するため，文字テンプレート内の文字種数については検索精度に影響は少ない．一方，文字認識手法においては文字テンプレート内の文字種数の変化が文字認識精度に影響を与え，検索精度が低下すると考えていたのだが，検索結果にほとんど差はみられなかった．一般に，文字認識に用いられる文字テンプレートには数万種単位の文字が含まれており，類似形状の文字が文字認識精度に影響を与える．今回の文字テンプレート内には 1,000 種類程度の文字しか含まれておらず，類似形状の文字が文字テンプレート内に少なかったため，検索精度に差が現れなかったと考えられる．今後，大規模な文字テンプレートを用いた検証が必要である．

次に，各手法における検索精度について考察する．1-gram 特徴量を用いた検索結果と 2-gram 特徴量を用いた検索結果を比較すると，2-gram 特徴量を用いた方がどの文字テンプレートを用いた場合においても検索精度が向上した．これは，2-gram 特徴量の場合は n 文字目の特徴量と $n+1$ 文字目の特徴量を合わせて特徴量照合することにより，隣接する文字の関係がより鮮明になるためと考えられる．

また，1-gram 特徴量を用いた場合と文字認識手法を比較した場合，文字認識候補が上位 1 件の場合には 1-gram 特徴量を用いた方が高い精度を得ることができた．しかし，候補数を上位 2 位にした場合には文字認識手法の方が高い検索精度を得られた．ただし，文字認識手法では候補数をさらに増加させると文字の誤認識による過剰検索が発生し，適合率が低下した．特に，検索キーワードがひらがな，カタカナで構成される場合に過剰検出が多く発生し，適合率低下の原因となっていた．これは，漢字と比較してひらがな，カタカナは使用頻度が高く，また，類似した形状の文字が比較的多いことが誤認識の原因となり，正確なテキストが作成されず検索精度が低下したと考えられる．

また，字幕検索システムにおいてはユーザの検索する字幕をもれなく検索することが重要である．つまり，

検索結果に多少のノイズを含んでいても，検索結果の上位に正解となるすべての字幕が含まれるシーンを出力することが必要であり，そのためには，再現率を100% にすることが重要である．図 11 より，文字認識手法では再現率を向上させようとすると適合率が大きく低下した．さらに，文字認識候補数を上位 7 件まで増加させても再現率は 100% に達しなかった．このことから，文字認識手法では検索ノイズが多く，検索もれが発生する確率が高いといえる．たとえば，実験で使用した 91 単語を使用して検索される総字幕数は 811 件であるため，最も精度の良い文字認識候補上位 2 件の場合でも，約 8 件の検索もれが発生することになる．一方，提案手法では，再現率が 100% に達しており検索もれがない．さらに，適合率も高く検索ノイズが少ないため，字幕検索に有効である．

次に，各手法における前処理および検索時間について考察する．各手法の前処理の時間については，提案手法では前処理での特徴量照合が必要なく，テンプレートの作成時間のみを要する．一方，文字認識手法は文字認識の過程が必要となるため，より多くの時間を必要とする．各手法の検索時間を比較すると，文字認識手法の検索時間が提案手法に対して，かなり高速である．今回の実験では，特徴量照合のアルゴリズムは同一のものを使用している．このため，文字認識手法では前処理の段階で特徴量照合を行いテキスト化が完了しているのに対し，提案手法では検索過程で特徴量照合を行い 4.4 節で述べた検索結果の絞り込みを行っている．そのため検索時間に大きな差が現れたと考えられる．そこで，表 9 の各処理部における所要時間に着目する．キーワードの分割，特徴量作成，特徴量照合の各処理は高速であるが，字幕決定に非常に時間がかかっていることが分かった．これが，検索時間に大きな差が現れた原因であると考えられる．また，表 9 と表 10 を比較すると，処理時間のほとんどが字幕の決定過程で使用されていることが分かった．字幕決定時のデータ数の増加が影響したと考えられる．提案手法の検索時間を高速化するためには 4.4 節で示した字幕の決定方法を再検討する必要がある．現在の字幕決定アルゴリズムでは，検索結果内の各字幕番号と出現位置の情報を個別に照合して字幕特定条件に合うものを選別している．この方法では，検索キーワードが N 件からなり，1 文字分の検索結果に M 個の字幕情報を持つとすると，M^N 回の照合が必要になる．そこで，今後は，1 文字分の検索結果をビットベクトル化する等して，より少ない照合で最終的な検索結果を得られるような改良案を組み込みたい．

7. ま と め

本論文では，高精度かつ高速な字幕検索システムについて提案した．本研究では，字幕検索には映像内のすべての字幕を認識する必要はなく，検索キーワードの出現位置のみが認識できれば字幕検索が可能である点に着目した．検索キーワードに対応する特徴量だけを照合することで従来手法の問題点であった再現率の低下を改善し，検索精度を向上させた．実際の映像データを用いた検索実験では 1-gram 特徴量を用いた場合最大 98.61%，2-gram 特徴量に拡張することにより最大 99.59% の平均適合率を得ることができ，高精度の字幕検索が可能であった．また各手法の F 値を求めた結果，2-gram 特徴量を用いた提案手法が最も良い結果を得ることができた．さらに，検索時間に関しては 2-gram 特徴量を用いた場合でも 3 時間の映像から約 0.5 秒で検索結果を得ることができたが，字幕決定に多くの時間を要するため従来手法に比べて長い検索時間を必要とした．

今後の課題として，より大規模な映像データを用いた評価実験，背景ノイズのある字幕を用いた評価実験を行う必要がある．また字幕決定を高速化するため，各文字の検索結果の絞り込み手法，および，類似度の算出法について再検討する必要がある．また，特徴量に関しては 2-gram からさらに，n-gram に拡張し実験を行う予定である．

謝辞 本研究の一部は，科学研究費補助金基盤研究（B）（17300036），科学研究費補助金基盤研究（C）（17500644）を受けて行われた．

参 考 文 献

1) Milan, P. and Willem, J.: *Content-based video retrieval: A database perspective*, Kluwer Academic Publishers (2003).
2) 木村昭悟，柏野邦夫，黒住隆行，村瀬 洋：グローバルな枝刈りを導入した音や映像の高速検索，信学論（D-Ⅱ），Vol.J85-D-Ⅱ, No.10, pp.1552–1562 (2002).
3) Lienhart, R.: Automatic text recognition for video indexing, *Proc. 4th ACM Multimedia*, pp.11–20 (1996).
4) 新井啓之，桑野秀豪，倉掛正治，杉村利明：映像中のテロップ表示フレーム検出方法，信学論（D-Ⅱ），Vol.J83-D-Ⅱ, No.6, pp.1477–1486 (2000).
5) Mita, T. and Hori, O.: Improvement of video text recognition by character selection, *Proc. 6th ICDAR*, pp.1089–1093 (2001).
6) Gaede and Gunther, O.: Multidimensional ac-

cess methods, *CMComputing Surveys*, Vol.30, No.2, pp.170–231 (1998).
7) 佐藤　隆，新倉康巨，谷口行信，阿久津明人，外村佳伸，浜田　洋：MPEG 符号化映像からの高速テロップ領域検出法，信学論（D-Ⅱ），Vol.J84-D-Ⅱ, No.8, pp.1847–1855 (1998).
8) 堀　修，三田雄志：テロップ認識のための映像からのロバストな文字部抽出法，信学論（D-Ⅱ），Vol.J84-D-Ⅱ, No.8, pp.1800–1808 (2001).
9) 中嶋正臣，米倉雄司：平滑化周辺分布と判別分析を用いた手書き文字切出し方式，信学論（D-Ⅱ），Vol.J78-D-Ⅱ, No.7, pp.1039–1046 (1995).
10) 能隅進一，福田亮治，玉利文和，鈴木昌和：絞り込み法による数式文字認識とその日本語/数式領域切出しへの応用，信学論（D-Ⅱ），Vol.J83-D-Ⅱ, No.3, pp.895–906 (2000).
11) 堀桂太郎，根本孝一，伊藤彰義：文字の輪郭線に着目した手書き漢字の特徴抽出法－外郭局所的輪郭線特徴と外郭局所的モーメント特徴，信学論（D-Ⅱ），Vol.J82-D-Ⅱ, No.2, pp.188–195 (1999).
12) 森　稔，倉掛正治，杉村利明，塩　昭夫，鈴木章：背景･文字の形状特徴と動的修正識別関数を用いた映像中テロップ認識，信学論（D-Ⅱ），Vol.J83-D-Ⅱ, No.7, pp.1658–1666 (2000).
13) 森　稔，澤木美奈子，荻田紀博：特徴補正に基づくカテゴリー依存特徴抽出法による映像中文字認識，信学論（D-Ⅱ），Vol.J87-D-Ⅱ, No.8, pp.1632–1640 (2004).
14) 北　研二，津田和彦，獅々堀正幹：情報検索アルゴリズム，共立出版 (2002).
15) 味岡四郎，柘植　覚，獅々堀正幹，北　研二：順位キューを用いた多次元データの高速近傍検索アルゴリズム，電気学会論文誌 C，Vol.126, No.3, pp.353–360 (2006).

(平成 19 年 6 月 20 日受付)
(平成 19 年 10 月 16 日採録)

(担当編集委員　橋本 隆子)

西川　伸紀（学生会員）

平成 14 年徳島大学工学部知能情報工学科卒業．平成 17 年同大学大学院工学研究科博士前期課程知能情報工学専攻修了．現在同大学院工学研究科博士後期課程在学中．マルチメディア情報検索の研究に従事．

獅々堀正幹（正会員）

平成 3 年徳島大学工学部情報工学科卒業．平成 5 年同大学大学院博士前期課程修了．平成 7 年同大学院博士後期課程退学．同年同大学工学部知能情報工学科助手．平成 9 年同大学工学部知能情報工学科講師．平成 13 年同大学工学部知能情報工学科准教授．現在同大学工学部知能情報工学科助教授．博士（工学）．マルチメディア情報検索，自然言語処理の研究に従事．著書『情報検索アルゴリズム』(共立出版)，情報処理学会第 45 回全国大会奨励賞受賞．電子情報通信学会，言語処理学会各会員．

柘植　　覚（正会員）

平成 8 年徳島大学工学部知能情報工学科卒業．平成 10 年同大学大学院工学研究科博士前期課程知能情報工学専攻修了．平成 13 年同大学院工学研究科博士後期課程システム工学専攻修了．平成 12 年徳島大学工学部助手，現在，同大学工学部知能情報工学科講師．博士（工学）．音声認識，情報検索等の研究に従事．日本音響学会会員．

北　　研二（正会員）

昭和 56 年早稲田大学理工学部数学科卒業．昭和 58 年沖電気工業（株）入社．昭和 62 年 ATR 自動翻訳電話研究所出向．平成 4 年徳島大学工学部講師．平成 5 年同助教授．平成 12 年同教授．平成 14 年同大学高度情報化基盤センター教授．博士（工学）．自然言語処理，情報検索等の研究に従事．平成 6 年日本音響学会技術開発賞受賞．著書『確率的言語モデル』(東京大学出版会)，『情報検索アルゴリズム』(共立出版) 等．電子情報通信学会，言語処理学会各会員．