様式Ｃ－１９

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２４年　５月２５日現在

機関番号：１６１０１
研究種目：基盤研究(C)
研究期間：2009～2011
課題番号：２１５００９４０
研究課題名（和文）　教育用映像素材に対する字幕検索システムの開発

研究課題名（英文）　Development of the Telop Retrieval System for Educational Video Data

研究代表者
獅々堀　正幹（SHISHIBORI　MASAMI）
徳島大学・大学院ソシオテクノサイエンス研究部・教授
研究者番号：５０２７４２６２

研究成果の概要（和文）：本研究では映像内の字幕に着目した字幕検索システムを開発する．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．教育用映像データに対して検索実験（検索単語として 91 単語）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．

研究成果の概要（英文）：Video telop retrieval systems based on telop characters can retrieve the corresponding telops to the query from the huge video data. On this research, a new telop retrieval system based on telop characters is developed. In order to specify the suitable telop, this method calculates the distance between each image features of telop characters and template image features of query keyword. 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.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
２００９年度	1,500,000	450,000	1,950,000
２０１０年度	1,300,000	390,000	1,690,000
２０１１年度	500,000	150,000	650,000
総　計	3,300,000	990,000	4,290,000

研究分野：総合領域
科研費の分科・細目：科学教育・教育工学
キーワード：教材情報システム，字幕認識，字幕検索，映像検索，映像処理

１．研究開始当初の背景

現在，マルチメディア技術の発展に伴って，教育分野の中で教育用映像素材が頻繁に利用されつつある．大量の映像素材がアーカイブ化されつつある反面，映像題目，映像制作会社，映像ジャンルといった予め人手で付与されたキーワード情報を基に管理されているのが現状である．そのため，学習者が事前にキーワード情報について熟知していなければならず，また，予め決められたキーワード体系だけでは目的の教育用映像素材を絞り込むことができず，教育効果を低下させる原因になっている．そこで，より教育効果を上げるためには，映像内容に準じた情報で映像を検索する必要がある．映像内容を表すコンテンツとしては，映像フレームの構図情報，音声情報等が挙げられるが，本研究では，ユーザからの入力が容易なテキスト情報に対応した字幕情報に着目する．

本研究では，教育用映像素材の中でも特に字幕情報を有効利用できる映像ジャンルとして，英会話教育用の映像素材を対象にする．英会話教育用の映像素材には，テロップと呼ばれる字幕が付与されており，現在話題としている内容を的確に表している．このような字幕に基づいて映像シーンを検索できれば，大量の英会話映像素材集から特定の単語が英会話で使用されているシーンのみを集めて試聴することができる．つまり，学習者が詳しく知りたい単語についての用例を，あたかも辞書を調べるように，解説付きの映像データ集で学習することが可能になる．

２．研究の目的

本研究では，教育用映像素材の中でも特に英会話教育用の映像素材を対象にし，映像から字幕領域を検出する技術，検出した字幕領域から字幕文字を認識する技術，及び学習者が字幕により学習に有効な映像シーンを検索する技術を開発することを目的とする．

まず，映像データから字幕領域を検出する技術に関しては，一般にオープンキャプションには様々な形や色の字幕が用いられている点に着目する．そこで，字幕の大きさ，フォント，色彩に特化せず，かつ，背景の影響に頑健な字幕領域検出技術を開発する．また，映像内には静止した字幕だけでなく，スクロールする字幕も頻繁に出現する．そこで，スクロール字幕に対して，字幕出現時間と字幕領域の検出技術を開発する．

次に，字幕により映像シーンを検索する技術に関しては，字幕による映像シーン検索を実現するうえで，字幕領域を OCR 等で文字認識してテキスト情報に変換した後，テキストベースで映像を検索することが考えられる．しかし，文字の誤認識を発生した場合，誤認識した字幕を含む映像シーンは検索漏れを起こすことが予測される．そこで，検索漏れを極力抑制し，かつ，文字認識等の前処理を必要としない映像シーン検索技術を開発する．

最後に，教育効果が向上する字幕検索システムの利用方法の確立を目的として，英会話教育用映像素材に対して字幕検索システムを適用し，より教育効果が向上する字幕検索の応用システムを構築する．また，他の教育用映像素材に対しての適用方法を検討する．

３．研究の方法

(1) 字幕領域を検出する技術について

本論文では，文字候補画像と領域分割処理を施した領域分割画像とを用いた字幕領域の検出手法を提案する．以下，文字候補画像と領域分割画像とを用いて生成された字幕画像を文字領域画像と呼ぶ．まず，文字候補画像と領域分割画像とを用いて文字領域画像を生成する手法を説明した後，文字領域画像と原画像とを用いて文字色の自動設定を行い，最終的な字幕領域を検出する手法の流れを説明する．

領域分割画像を用いた字幕検出手法の流れを図１に示す．まず，従来手法と同様の方法で文字候補画像を得る．同時に，領域分割した画像（以下，領域分割画像）と領域分割画像データを得る（図１の step5）．領域分割画像データには，領域番号，各領域の大きさ（総画素数），領域の中心座標，領域の色情報，領域に属している座標が記載されている．なお，文字候補画像は連続した 4 フレームから作成しているが，領域分割画像は文字候補画像を作成する際に使用した最初の１フレームから作成している．

次に，これら 2 つの画像を用いてノイズを除去する（図１の step6）．ここでは，文字候補画像を水平方向に走査することで字幕部分の画素だけを残すノイズ除去，および文字候補画像と領域分割画像データとを照らし合わせて字幕らしい分割領域のみを選択するノイズ除去の２種類の処理を行う．その後，一般に字幕には縁取り部分が設定されているため，縁取り部分を除去する処理を行う（図１の step7）．具体的には，字幕本体と縁取り部分は異なる文字色が使用されているため，ノイズ除去の段階で残存している白画素が属している領域の色を k-means 法によりクラス分けを行う．最後に，再現率を向上させるため，文字の補間を行う（図１の step8）．具体的には，step7 までで字幕として残っている画素周辺の各分割領域を探索し，領域の大きさや代表色が字幕領域と類似している領域を増殖させる．図２に文字領域画像の検出例を示す．

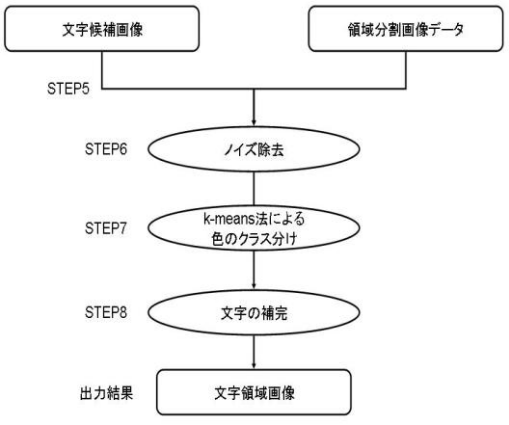

図１　領域分割画像を用いた字幕検出手法の流れ

図２　文字領域画像の検出例

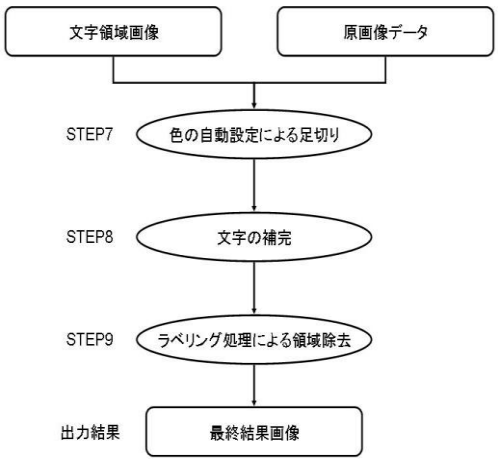

図３　字幕文字色を用いた字幕検出手法の流れ

図４　最終的な文字領域画像の検出例

前述の方法で生成した文字領域画像は，字幕部分に存在する微小な分割領域に対して検出漏れを起こす傾向があり，再現性が低下する特徴がある．そこで，文字色に着目して再現率を向上させる手法を適用する（図３）．まず，連続フレーム間で生成した複数の文字領域画像と原画像とを用いて字幕の色情報を特定する（図３の step7）．その後，文字の補完を行い（図３の step8），最終的に残った画素に対してラベリング処理を行い，大き過ぎる領域を削除する（図３の step9）．図４に最終的な文字領域画像の検出例を示す．

(2) 字幕を検索する技術について

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．

字幕検索手法の概要を図５に示し，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：字幕領域の切り出し**
  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．
- **前処理手順２：文字画像の切り出し**
  字幕画像を２値化して文字列画像を切り出し，文字列画像から文字画像を切り出す．
- **前処理手順３：特徴量抽出**
  切りだした文字画像から特徴量を抽出する．
- **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：検索キーの特徴量の取得**
  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．

［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高々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．

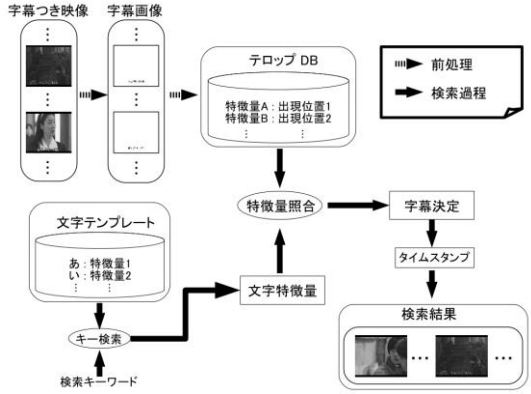

図５　字幕検索手法の概要図

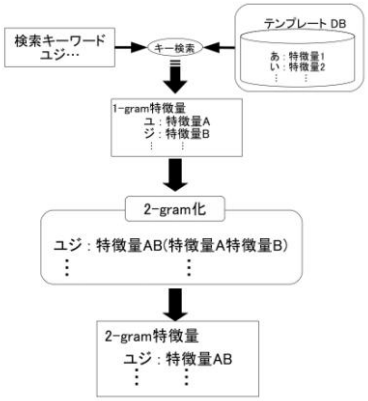

図６　2-gram 化の手順

上記で述べた手法では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問題点があった．更に，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．検索精度を更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，１文字分の特徴量を 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 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６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 の検索手順４

［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で 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 特徴量に拡張することが可能である．

４．研究成果

(1) 字幕領域検出技術に対する評価

提案手法の有効性を確かめるために実験を行った．実験データとして，RGB フルカラー，解像度 352×240，フレームレート 29.97fps の静止字幕が出現するドラマ映像を用いた．正解データとして，このドラマのオーバーレイ領域の字幕のみの画像を使用した．正解データと文字候補画像，文字領域画像を画素毎に照合し，適合率，再現率を算出した．評価する画像はドラマ中から無作為に選んだ 30 枚とした．また，検出単位は画素単位で数える．字幕が表示されているフレームにおいて，字幕の表示位置の画素を検出できていれば正検出，その他は検出漏れとした．字幕が表示されていないフレームや字幕が表示されていない位置の画素を検出した場合には誤検出とした．従来手法（文字候補画像）と提案手法（文字領域画像と最終結果画像）との比較実験結果を図７に示す．

図７より，従来手法に比べ，提案手法は適合率，再現率ともに上昇した．まず，適合率が上昇した要因として，ノイズ除去が挙げられる．文字候補画像は，動きの少ないシーンの場合，建物などが写っていると字幕以外にも画素が多数残ってしまい，適合率が低かった．しかし，提案した手法により字幕以外の部分を取り除くことができ，適合率が上昇したと考えられる．また，字幕部分においても，文字候補画像はエッジの部分が強く残る傾向があり，文字の縁取りだけが残っている場合が多かった．それを今回の手法で文字の欠損部分を補完できたのも適合率の上昇に関係していると考えられる．

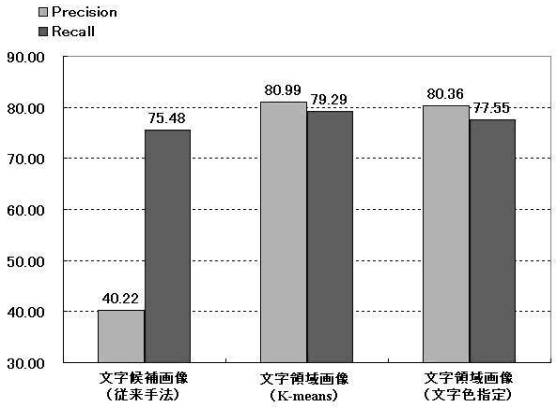

図７　字幕領域検出手法の実験結果

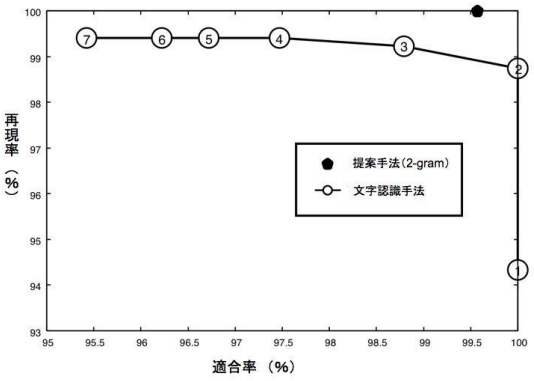

図８　従来手法と提案手法との精度比較

しかし，領域分割画像を用いた手法の再現率は，従来手法に比べてあまり精度が上がらなかった．この点について，今回の提案手法では，k-means 法で色のクラス分けをする際に，単純に各クラスの要素数を基準にしてクラスを選択している．そのため，縁取りの画素が多く残ってしまっていると，字幕部分の補完がほとんどできなくなり，再現率の低下をまねいたと考えられる．今後は，クラス内の各要素の位置も考慮し，より精度よく文字本体のクラスを選択できるアルゴリズムを考案する必要がある．

また，提案手法の領域分割画像を用いた手法と，その後に文字色の自動設定を適用した手法とを比べると，文字色の自動設定を適用した方が精度の向上が見受けられた．このように領域分割画像を用いた手法だけでは対応できない字幕に対しても，文字色も考慮することで改善でき，再現率が上昇した要因だと考えられる．

(2) 字幕検索技術に対する評価

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，及び 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 話分ずつ増加した 9 種類のテンプレートを作成した．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付けが必要であるため，書きおこしテキストを用いて対応付けを人手で確認して作成した．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．

表１　従来手法と提案手法との検索時間比較

	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

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，及び検索時間を比較した．更に，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．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 3.20GHz
・ メモリ PC-4200 DDR2 SDRAM 1G byte

② 実験結果及び考察

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８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１に示す．ただし，文字認識手法の F 値は最も精度のよ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及び検索の所要時間を表１に示す．

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８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．更に，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．例えば，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，最も精度のよ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．更に，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及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検索結果のしぼりこみを行っている．そのため検索時間に大きな差が現れたと考えられる．

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３件）

① M.Shishibori, S.Lee, K.Kita, An E-mail Filtering Method Based on Multi-attribute Values of User's Profile, International Journal of Computer Applications in Technology, 査読有, Vol.40, No.4, 2011, pp.273-279, DOI:10.1504/IJCAT.2011.041656
② M.Shishibori, D.Koizumi, K.Kita, Fast Retrieval Algorithm for Earth Mover's Distance Using EMD Lower Bounds and Skipping Algorithm, International Journal of Advances in Multimedia, 査読有, Article ID 421820, 2011, pp.1-9 DOI:10.1155/2011/421820
③ Y.Matsumoto, T.Uemiya, M.Shishibori, K.Kita, A Method for Detecting Subtitle Regions in Videos Using Video Text Candidate Images and Color Segmentation Images, International Journal of Advanced Intelligence, 査読有, Vol.2, No.1, 2010, pp.37-55 http://aia-i.com/ijai/sample/vol2/no1/37-55.pdf

〔学会発表〕（計５件）

① M.Shishibori, An Improved Method to Select Candidates on Metric Index VP-tree, Proc. of the International Conference on Knowledge Discovery and Information Retrieval, 2011.10.15, Paris, France
② 丸西立起，SIFT 特徴量を用いた映像データに対する人物検索システムの開発，第３回データ工学と情報マネジメントに関するフォーラム，2011.3.1，ラフォーレ修善寺（静岡県）
③ 西野裕一，Earth Mover's Distance を用いたSIFT特徴量に基づく類似画像検索手法，第 3 回データ工学と情報マネジメントに関するフォーラム，2011.3.1，ラフォーレ修善寺（静岡県）
④ M.Shishibori, A Method to Improve Metric Index VP-tree for Multimedia Databases, Proc. of 17th Korea-Japan Joint Workshop on Frontiers of Computer Vision, 2011.2.1, Ulsan, Korea
⑤ 大川文也，映像データからの字幕領域の検出手法，平成21年度電気関係学会四国支部連合大会，2009.9.25，愛媛大学（愛媛県）

６．研究組織

(1)研究代表者
獅々堀　正幹（SHISHIBORI MASAMI）
徳島大学・大学院ソシオテクノサイエンス研究部・教授
研究者番号：５０２７４２６２
(2)研究分担者
北　研二（KITA KENJI）
徳島大学・大学院ソシオテクノサイエンス研究部・教授
研究者番号：１０２４３７３４
柘植　覚（TSUGE SATORU）
大同大学・情報学部・准教授
研究者番号：００３２５２５０
(3)連携研究者
（　　）
研究者番号：