ユーザによる動き方と位置の指定に基づいた技能映像インデクシング手法

宮下広夢† 嶌田聡† 東野豪†

実技映像をすぐさま振り返ることで，技能伝承の効率を高めることができる．しかし映像は扱いが難しく，撮影が面倒であったり，見たいシーンを探しづらいといった問題がある．特に技能伝承の分野では映像を作業の試行や手順に分けて扱いたいというニーズがあった．そこで本研究ではユーザによる動き方と位置の指定に基づいた技能映像インデクシング手法を提案する．本提案では被写体の動き方と位置に試行や手順の区切りに対応した規則性があることに着目し，ユーザが指定した動きを映像全体から検索しインデクシングする．また動きの軌跡やインデクシングの結果をユーザにフィードバックし，インデクスの効率的な修正を実現した．

An Indexing Method for Skill Training Videos based on Clipping Motion and Area by the User

Hiromu Miyashita,† Satoshi Shimada† and Suguru Higashino †

Reviewing of skill practice videos can enhance efficiency of skill trainings. However, videos are hard to deal, laborious to capture, and difficult to search needed scenes. In various fields of skill tradition, there is a compelling need for easy access to each trials and procedures in videos. In this paper, we propose an indexing method for skill training videos based on clipping motions and areas by the user. When a user clips a motion that is a delimiter of trials and procedures, indexes are added on some scenes that have similar motions. Additionally, we implemented visualizations of motion trajectories and feedbacks of indexing results for remediations of indexings.

1. はじめに

医療看護やスポーツ，技芸の世界では己の技能を高めることが重要である．熟練者や自らの動きを映像で蓄積することで，指導教育や自己学習による技能の洗練の効果を効率的に高めることができる．そこで我々は映像を活用した技能の蓄積，伝承，発見について取り組んでいる．この取り組みでは，映像活用型ナレッジ発見・共有システム[1] や，シーンを集約して俯瞰的な議論を促すようなシステム[2] を提案した．

本研究では特に現場指導における映像活用に注目し，指導者や生徒がその場で実技を振り返ることで，実技の感覚が新鮮なうちに内省を促すようなシステムを目指している．そのためには普段の実務や演習を撮影する必要があり，指導者は演習を滞らせないように見せたいシーンをすぐさま提示する必要がある．

しかし実際には撮影は面倒であり，実技や演習に集中したいユーザにとって試行ごとの撮影 ON/OFF 操作や手動のシーン定義は非常に面倒である．撮りっぱなしにすると不要な区間も撮影されてしまう上，試行や手順の区切りが不明確になってしまう．すると実技を作業ごとに確認したい，複数の試行を見比べたいといったときに，映像を一から見直さなければならない．

そこで本稿ではシーンの選択や引用を簡単にするために，未編集の実技映像から試行や手順の区切りとなるようなシーンを検出する映像インデクシング手法を提案する．

2. インデクシングする映像の特徴と課題

2.1 撮影の方針と映像の特徴

本研究では普段の実技や演習を撮影し，その場で映像を振り返ることを想定している．実技の進行を妨げないように，ユーザにかかる負担はできるだけ軽減する必要がある．特に撮影のために実技の内容に手を加えることは本末転倒であり，技能の鍛錬を阻害してしまうと考えている．

またビデオカメラの扱いについても同様であり，で

† NTT サイバーソリューション研究所
NTT Cyber Solutions Laboratories

きるだけ操作を簡易にしなければユーザにとって負担となってしまう．そのため本研究では撮影は固定カメラによる撮りっぱなしとする．撮影の場所や状況，対象とする技能によってビデオカメラの設置は任意に調整できる必要がある．各実技や演習を開始する前にカメラの設置を完了してしまえば，後はカメラを気にしなくとも実技に専念できる，ということが理想的である．

2.2 インデクスを付ける対象

人物の動作はいくつかの手順にまとめることができ，一連の手順の流れが複数回繰り返される場合は，試行が連なっている，と捉えられる．実際の現場指導でも実技を試行や手順のまとまりとして扱っている．例えばサーブ練習を何回も実施する場合，指導者は「前のサーブと比較したい」など試行を単位として指導することが一般的である．またマニュアルを作成する場合，ある作業をいくつかの手順に分けて順に注意点やコツをまとめると，参照する人にとっても分りやすくなる．そこで本研究では技能映像を扱いやすくするために，試行と手順の区切りについてインデクスを付加することを目指す．

2.3 従来技術と課題

技能的な動作を対象として，被写体の動きに基づいて撮りっぱなしの映像をインデクシングする手法としては，野球の投球イベント検出[3]がある．これは打捕球音と $MotionHistoryImage(MHI)$ を特徴量とした教師なしクラスタリングを行っている．しかしこの手法は実技中に特定の音響変化が必ず発生するような技能分野にしか適用できない．

他に監視カメラ映像から非定常行動を検出する手法[4]や，$Pointing, CelltoEar, ObjectPut$ といった動きを検出する手法[5]がある．これらは機械学習を利用して，いつもと違う動きや特定の行動をクラスタリングによって検出する手法である．本研究では撮影環境が毎回異なることから事前の教師データ取得が困難であり，試行や手順の区切りとなる動きは技能の分野によって異なるため機械学習では検出できない．

一方で技能映像からユーザの見たいシーンを見せるための研究としては，サッカーのシーンを検索する手法[6]がある．これはサッカーの選手やボールの移動軌跡を GUI で画面上に直接描画させ，オブジェクト位置の時系列データから検索している．ただし時系列データはセンサによる取得を前提としており，本研究では利用できない．

よって，理想とする映像活用のためには特定の行動検出や機械学習によらない手法が必要である．

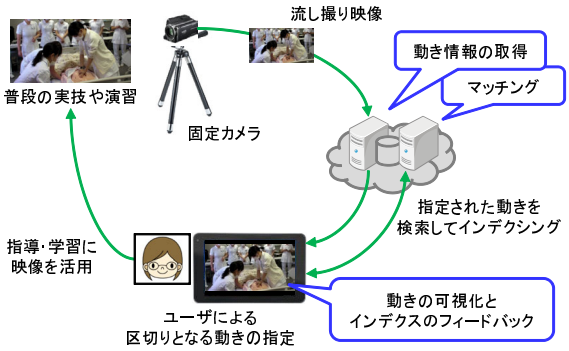

図 1 実技映像インデクシングの流れ
Fig. 1 Indexing flow for a skill learning video

3. 技能映像のためのインデクシング手法

3.1 技能映像に含まれる動きの規則性

目標とするような映像インデクシングを実現するためには，試行や手順の区切りで発生し，かつ映像から取得できる何らか特徴を見つけ出す必要がある．そこで我々は静脈注射や移乗介護のような医療看護の映像や，スポーツ，趣味技芸などの映像を多数視聴し，試行や手順の区切りと対応する規則性について検討した．

例えば静脈注射の様子を俯瞰した映像の場合，消毒，注射，止血など各手順の合間には必ずトレーから何かを持ち上げ，患者の腕へ持っていく動きが発生する．またテニスやバレーなどではサーブをラリーの区切りとおくことができるが，そのサーブはすべて特定のライン上で実施される．

この検討から，様々な技能映像の傾向として被写体の動き方とその位置が試行や手順の区切りに対応した規則性があることがわかった．ただし映像中のどの動き方と位置が実技の区切りと対応しているかは撮影内容毎に異なり，また構図によって映像上の見え方も変化するため一意に定めることができない．

3.2 ユーザによる動き方と位置の指定に基づいた技能映像インデクシング手法

本研究テーマでは普段の実技や演習を撮影し，その場で視聴することを想定している．そのためユーザはその技能や演習の流れを前提知識として持っており，試行や手順の区切りと対応する動きや，区切りの数や分布をある程度把握していると考えられる．さらにカメラは固定のため映像中に画角や環境が大きく変化することはなく，動作の場所が同じであれば映像画面上の同じ位置に描画される．

そこで本研究では試行や手順の区切りと対応する動き方と位置の特定を，ユーザに指定してもらうことを提案する．提案手法の流れを**図 1** に示す．撮影した映

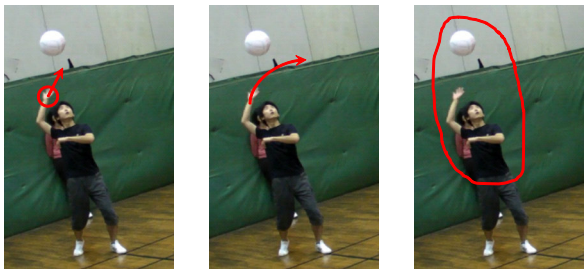

図 2　動き方と位置の指定方法（左から点指定，線指定，領域指定）
Fig. 2　Marking methods for a motion and position (point marking, line marking, area marking)

像をユーザに提示し，区切りと思われる動きをユーザに指定してもらう．指定された動きをクエリとして映像全体から似た動きを検出することで，映像中の区切りにインデクスを付加できる．

またインデクスの結果をユーザに分りやすく提示することで，インデクスの間違いや過不足をユーザに気付かせることが可能だと考えた．そこで本研究ではインデクシングのフィードバックを行い，ユーザに修正を促すことを提案する．

3.3　動きの指定方法の検討

ユーザによるクエリの作成については，GUI による映像画面上への直接描画が適当であると考えた．文字や音声では動きと単語を対応付けることが困難であり，映像を視聴しながら動きを確認して直接指定するほうが直感的なためである．GUI による直接描画として**図 2** のような手法を実装し，検討を行った．

点指定: 画面上のある位置を指定し，そこから動いた方向を指定する．指定座標やその付近の動きを参照するだけのため，細かな動きの違いなどを区別できず，誤ってシーンを検出してしまうことが多い．

線指定: 画面上をなぞった軌跡付近の動きから位置と動き方を指定する．点指定と比較して被写体の移動方向が途中で変化する場合でも検索しやすい．ただし動きの軌跡とずれてしまうと検出できなくなる，複雑な動きや往復するような動きは線が重畳してしまうなどの問題が発生する．

領域指定: 指定したい動きのシーンで映像を一時停止し，動きが発生している領域を自由閉曲線で囲むことで，その領域内の前後 1 秒間の動きを例示的に指定する．人物の動作は一つの動きだけで表せることは少ないが，領域指定であれば一回の動作に含まれる複数の動きをまとめて指定できる．また画面に映っている動きを囲むだけで動き方と位置が取得できるため，ユーザは直感的かつ具体的に動きを指定できる．

以上の検討から，本研究では領域指定による動き方と位置の指定を採用する．

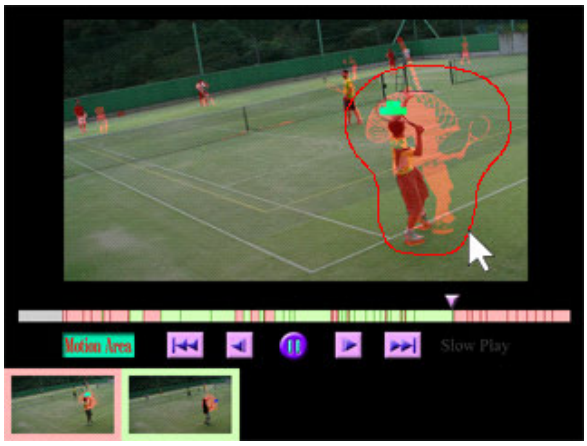

図 3　CliMISβ の画面構成図
Fig. 3　System diagram of CliMISβ

4. 実　　装

提案手法を実装したシステムを**図 3** に示す．以降，本システムは CliMISβ(Clipped Motion Indexing System β) と呼ぶ．なお，CliMISβ は映像インデクシングのテスト用であるため，撮影との併用はできない．

CliMISβ は映像中の動きに関する特徴量を取得し，ユーザが指定した動き方と位置に基づいてインデクシングを行う．この処理の流れを 4.1 で説明する．またシステムのユーザインタフェースについて，インデクシングの精度を高める工夫を施している．これについては 4.2 に詳しく述べる．

4.1　画像処理: 動き情報の取得とマッチング

本研究では映像に含まれる動きの情報としてオプティカルフロー (以下，フロー) を採用し，ブロックマッチング法で取得することとした．これは映像中の各フレームについて前フレームとの比較を行い，比較間隔ごとの各位置における動きの量と方向を計算する手法である．フローは MHI と比べて体や四肢のひねり，回転，オクルージョンなどの要素も表現できる．フローを取得してからインデクスを付けるまでの画像処理の流れを**図 4** に示し，以下に詳しく解説する．

(1)　フローの取得: 映像フレームが追加されるたびにフローを取得する．ベクトルの大きさで閾値処理を行い，大きいものだけ動きの角度を計算しメモリ上の配列に保存する．以降，この配列を動き配列と呼ぶ．

(2)　ROI の設定: クエリとしたい動きを，その動きが映っているフレーム上でユーザに囲んでもらう．その自由閉曲線の軌跡から，囲まれた領域の中に対応する動き配列の要素の座標を計算する．以降はこの領域内のフローを ROI フロー

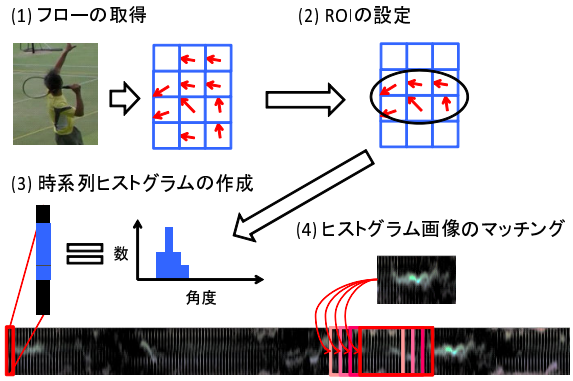

図 4　動き特徴量の処理の流れ
Fig. 4　Operations flow of motions features

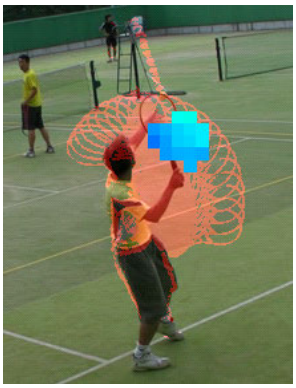

図 5　フローカラーブロックと動きシルエットの描画
Fig. 5　Flow color blocks and motion silhouettes

と呼ぶ．

(3)　時系列ヒストグラムの作成: 映像全体の各フレームについて ROI フローを参照し，横幅が 1pixel，縦幅が 24pixel のヒストグラム画像を作成して横に連結する．ヒストグラム画像の縦軸は角度であり，0 度から 30 度，15 度から 45 度，中略，345 度から 15 度といったように 15 度ずつ重畳のある 24 つにまとめている．またヒストグラムのパワーは ROI フロー中のフローの数であり，画素値の濃さで表現する．

(4)　ヒストグラム画像のマッチング: 時系列ヒストグラムから再生フレームの前後 1 秒分をコピーし，テンプレート画像とする．このテンプレート画像と似ている区間を映像全体のヒストグラム画像から探し出す．この結果は各フレームごとの類似度として得られ，類似度が閾値以上であればそのフレームの前後でクエリと同様の動きが発生したと判定してインデクスを付ける．ただし非常に近い位置にインデクスが付いた場合は一つにまとめる．

4.2　ユーザインタフェース: ユーザの判断を助ける

4.2.1　動きの可視化:領域指定をガイドする

本研究ではユーザによる動きの指定を簡単にするために，以下のような動きの可視化を行った．

フローカラーブロック: 4.1 で取得したフローに基づいて，画面上に色つきのブロックを描画する．描画するブロックの大きさと位置はブロックマッチング法で利用している値と同じであり，どの区画でフローが発生しているかを表示している．色は HSV 色空間を使ってフローの角度で決定し，同じような方向のフローは似た色で表示される．

動きシルエット: 再生しているフレームの未来と過去でどのような軌跡を描いて動きが発生してるのか，**図 5** のように可視化する．これは連続した 2 枚のフレーム画像の差分を計算した 1 枚の 2 値画像を作成し，画素値に差があったとみなせる場所について再生画面上に赤色のカラーリングを行っている．再生フレームより先のフレームを取得することで，動いた軌跡だけではなく，現在映っている被写体がその後どのように動くかといった未来の動きをも表示できる．

4.2.2　インデクスのフィードバック:修正を促す

また我々はインデクシングした結果をユーザへ 3 つの方法でフィードバックし，より理想的なインデクシングに近づけるようにインデクスを追加・修正できるようにした．

インデクスドシークバー: 映像のシークバー上のインデクスが付いたフレームと対応する場所にカラーリングや線の描画を行う．これにより映像全体のうちどのあたりにインデクスが付いたのかを一目で把握できる．例えばユーザがインデクスは映像全体に均等に付加されると判断していたときにインデクスの粗密を可視化することで，疎なところには検出漏れがある，もしくは密な区間には誤ってインデクスが付いている，といった気づきを与えることができる．

インデクス頭出し再生ボタン: 今再生しているフレームより前，後ろのインデクスを頭出しで再生できるような機能を持つボタンを付加する．これによりそれぞれのインデクスがユーザにとって正しいフレームに付加されているのかをスムーズに確認できる．このときユーザの想定とは違ったフレームにインデクスが付加されているのであれば，映像と可視化されたフローを眺めることで次にどのようにクエリを修正すればいいのかといった気づきを与えることができる．

クエリサムネイル: クエリを追加・修正するときに以前どのようなクエリを作成したのかを簡単に振り返れるように，クエリを作成したときのフレームのサ

図 6　実験で使用したテニスと陶芸の映像
Fig. 6　Videos of tennis and ceramic art

ムネイルを表示しておく．このときサムネイルの枠とシークバー上に描画するインデクスの位置を示す区間や線とを配色で対応付けることで，どのクエリによってどのインデクスが付加されたのかを明示できる．このフィードバックによってユーザは画面上の領域の囲み方とインデクスの付き方を内省し，より効率的なクエリの作成を実施できると考えらえる．

5. 実　　験

実装した CliMISβ を使用して，インデクシングの精度とフィードバックの効果を確かめた．被験者は 24〜27 歳の男女 4 名である．

5.1　実 験 方 法

実験では**図 6** に示すテニスのサーブ練習の映像と，電動ろくろの陶芸の映像について，被験者にインデクシングしてもらった．さらに被験者はインデクシングの結果を見て，必要と判断した場合のみ修正を行った．試行の区切りと見なせる動きは結果を比較するためこちらで設定し，テニス映像ではサーブ動作を，陶芸映像では手に水をつける動作を被験者に指定させた．なお被験者の作業画面は被験者用・実験者用の両方のディスプレイに映し出し，各操作でどのような傾向があるのかを確認できるようにした．

5.2　実 験 結 果

テニスのサーブ練習

図**図 7** に結果を示す．サーブ動作は 15 分の映像中に 53 回含まれている．すべての被験者において検出漏れが多数発生している．検出漏れの原因を調査したところ，被験者が指定した領域外でサーブを行っていたケースと，他人とサーブのクセが違ったケースに分けられた．フィードバック後はすべての被験者がインデクスの追加を行っており，その結果インデクシングの精度を大幅に改善している．

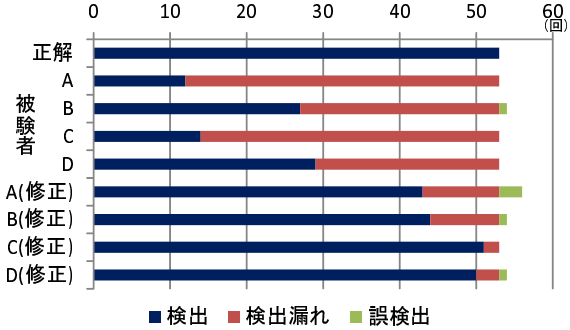

図 7　テニス映像のインデクシング結果
Fig. 7　Results of indexing for a tennis video

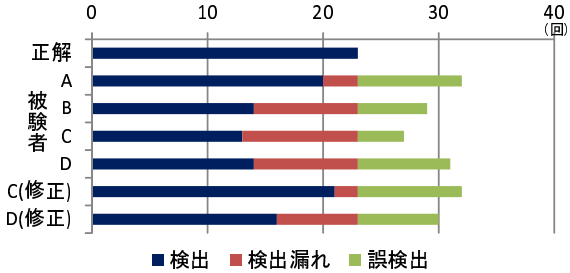

図 8　陶芸映像のインデクシング結果
Fig. 8　Results of indexing for a ceramic art video

電動ろくろの陶芸

図**図 8** に結果を示す．手に水を付ける動作は 32 分の映像中に 23 回発生していた．初回のインデクシングでは検出漏れと誤検出が同程度に多く，平均した再現率は 66%，適合率は 70%であった．主に手に水を付けるついでに洗っているシーンが検出から漏れており，画面右から講師が現れて作業の手伝いを行うシーンが誤って検出されていた．被験者 C と D がインデクスの修正を行っており，特に被験者 C は正しい検出の数を 13 から 21 へと大幅に改善させている．

6. 考　　察

実験の結果を踏まえ，インデクシングの精度に影響を与える要因とフィードバックの効果を検討する．

インデクシングの精度に与える要因: 指定したい動きの近くに別の動きがあるとクエリの中にノイズを含んでしまうため，インデクシングの精度が下がってしまう．現在は 2 秒分の ROI フローをもとに検索しているため，動きが切れ目なく連続的に遷移するなどの時間的に近い動きもノイズとなりやすい．また同じ動きでも作業者や試行によってクセがあると，検出漏れが多くなってしまう傾向があった．

しかし実験では被験者が指定したい動きを囲むとき，ノイズとなる動きをうまく除外できたケースや，クセ

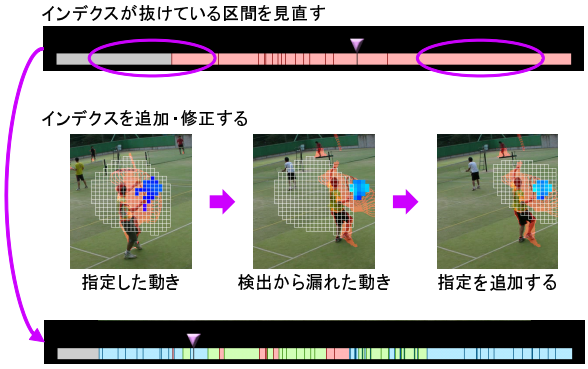

図 9　フィードバックによるインデクスの修正
Fig. 9　Remediating indexes with feedbacks

の現れにくい部分だけを囲むケースがあった．他にも同じ動きを対象としても被験者ごとに動きの囲み方が異なる傾向があり，インデクシングの精度も異なっていた．

フィードバックの効果:

被験者の操作を観察したところ，動きの可視化は動きを囲むときの形状やノイズとなる動きを区別する境界のガイドとして役割を果たしていた．特に動きシルエットは動きの発生する 1 秒前に予告的に描画されるため，ちょうど動きが行われているフレームを指定しやすくなる効果もあった．

またテニス映像に対する被験者 C のインデクシングとフィードバック後の修正の概要を**図 9** に示す．一回目のインデクシングのあと，この被験者はインデクスドシークバーを見てインデクスに粗密があることに気づいていた．疎の区間を改めて見直すことで検出漏れとなった試行を発見し，その動きをクエリとしてインデクスの修正を行っていた．このようにフィードバックによってインデクスの間違いをユーザに気付かせ，動きの可視化が具体的にどう修正すべきかのガイドを行うといった連携効果があったと考えられる．

7. おわりに

本稿では技能伝承のための映像インデクシング手法として，固定カメラの撮りっぱなし映像から実技の試行や手順の区切りをインデクシングすることを目標とした．医療看護やスポーツなどの技能映像を調査した結果，被写体の動き方とその位置に試行や手順の区切りと対応した規則性があることを発見した．区切りとなる動きはユーザにとっては自明なため，我々はユーザによる動き方と位置の指定を取り入れることでインデクシングが容易になると考えた．

提案システムではユーザによる指定をクエリとしてオプティカルフローの時系列ヒストグラムのマッチングにより類似した動き方と位置を検索する．また未来と過去の動きの軌跡やインデクスが付加されたフレームなど，インデクスの精度を上げるためのフィードバックを実装した．

実験ではインデクシングの精度とフィードバックの効果を確認した．その結果，ノイズとなる動きや動きのクセがインデクシングの精度を低下させるが，ユーザの囲み方によってそれらを回避できることがわかった．また動きの可視化とフィードバックにより，インデクスの間違いと改善についてユーザに気付きを与え，修正を促せることがわかった．

本システムはノイズとなる動きや動きのクセが強い場合はインデクシングの精度が低下してしまうことが課題である．今後は動きが連続的に行われるような複雑な技能を対象として検討を進めていきたい．

参　考　文　献

1) 東 正造, 寺中 晶郁, 嶌田 聡, 小島 明: 映像中のナレッジ発見・共有を想定したシステムの開発, 映像情報メディア学会技術報告, vol.32, No.38, pp.99-104, 2008.
2) 宮下 広夢, 嶌田 聡, 小島 明: 複数の映像シーンに基づくノウハウ発見・共有支援手法の検討, 電子情報通信学会技術研究報告, Vol.110, No.453, pp.281-286, 2011.
3) 三上 弾, 紺谷 精一, 森本 正志: 突発音検出と教師なし動きクラスタリングを用いた野球映像からの投球イベント検出, 電子情報通信学会論文誌 (D), Vol.90, No.2, pp.526-534, 2007.
4) 数藤 恭子, 小嶋 宏之, 磯 和之, 秦泉寺 久美, 大澤 達哉, 小池 秀樹, 仲澤 斉: 人物の位置と動き方向の学習に基づくリアルタイム非定常検出, 電子情報通信学会技術研究報告, Vol.108, No.94, pp.19-24, 2008.
5) Yoshihiko Kawai, Masaki Takahashi, Mahito Fujii, Masahide Naemura, Shinichi Sato: NHK STRL at TRECVID 2010: Semantic Indexing and Surveillance Event Detection, TRECVID 2010 Participant Notebook Papers and Slides, 2010.
6) 矢島 史, 角谷 和俊, 田中 克己: 映像上での動きの直接描画によるサッカー映像検索, 情報処理学会研究報告, データベース・システム研究会報告, vol.2002, No.41, pp.33-40, 2002.