機械学習を用いた任意背景リアルタイム被写体抽出技術

かきぬま　ひろかず　　ながお　じろう
柿沼　弘員　/長尾　慈郎
みやした　ひろむ　　とのむら　よしひで
宮下　広夢　/外村　喜秀
ながた　ひでのぶ　　ひだか　こうた
長田　秀信　/日高　浩太
NTTサービスエボリューション研究所

NTTサービスエボリューション研究所では，遠隔のスポーツ選手があたかも目の前にいるかのように感じさせる超高臨場感通信技術Kirari!の研究開発を進めています．本稿では擬似3D映像などを用いて選手の存在感までも遠隔地に提示する際に必須となる被写体抽出技術において，任意背景からリアルタイムに被写体を抽出するシステムを紹介します．

はじめに

映像中から人や物体の正確な領域を把握することは，高品質な画像編集や合成を実施するうえで必須の技術のため，コンピュータビジョンにおける主要な研究テーマの１つです．超高臨場感通信技術Kirari!においても，正確な被写体領域を抽出することは擬似3D映像提示を行い，高い臨場感を得るために必須となります．NTTサービスエボリューション研究所では，任意背景リアルタイム被写体抽出技術[(1)]において，グリーンバックなどのスタジオ設備を用いずに試合会場や演技している舞台映像からリアルタイムに被写体の領域のみを抽出するシステムを提案しています．本稿では，そのシステムを①従来では判別できなかったよりわずかな特徴量の差までも判別する機械学習を導入すること，②赤外線光などを用い抽出したいオブジェクトの特徴量を生成すること，でより正確に被写体領域を抽出できるようにシステム開発しましたので紹介します．

機械学習を用いたリアルタイム被写体抽出フレームワーク

リアルタイムに対象領域を抽出する処理には背景差分法が知られています．背景差分法は，入力画像と背景画像との差分をとり，しきい値処理することで変化のあった被写体の領域として抽出するため，高速な処理ができ，事前の準備が少ないことから広く用いられてきました．しかし，適切なしきい値決定は難しく，また，背景の変化などに対応できないなどの問題がありました．そこでNN（Neural Network: ニューラルネットワーク）を用い，入力される特徴ベクトルを別の特徴空間に変換し，識別を行う被写体抽出方法を開発しました．NNを用いることで，識別のための特徴空間はあらかじめ与えられる教師データによりNN内で決定され，より適切な特徴空間に自動で変換されることが期待されます．また，識別に用いる入力も，対象の画像だけでなく，リファレンスとなる被写体の特徴を有した画像や，時間の異なる画像や，領域情報，赤外線映像などを入れても，NN内で適切に特徴空間に変換され，背景や被写体の高次な特徴量を用いた被写体抽出が実行できるため，背景の変化などにロバストになることが期待されます．

開発システムのワークフローを**図１**に示します．開発システムは２段階の被写体抽出を行います．１段目で

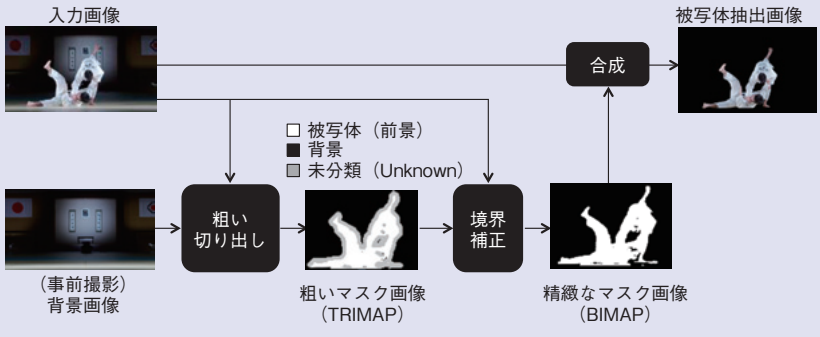
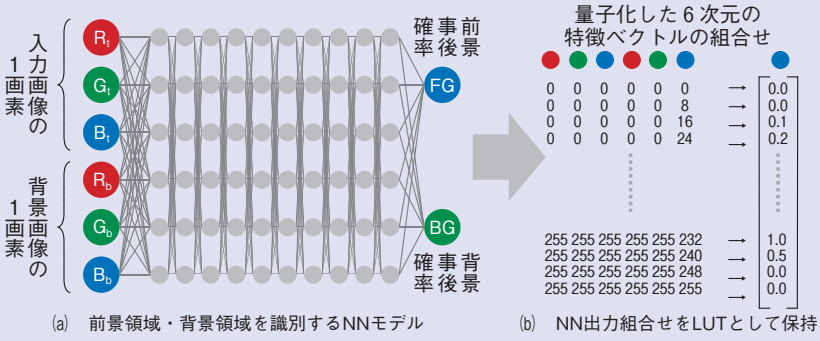


図１　リアルタイム被写体抽出フレームワーク

は，被写体領域を粗いマスク画像（TRIMAP*1）として抽出し，2段目ではTRIMAPを頼りにしたマッティング処理*2にて正確な被写体領域を抽出します．機械学習は1段目のTRIMAPを導出するのに導入しました．

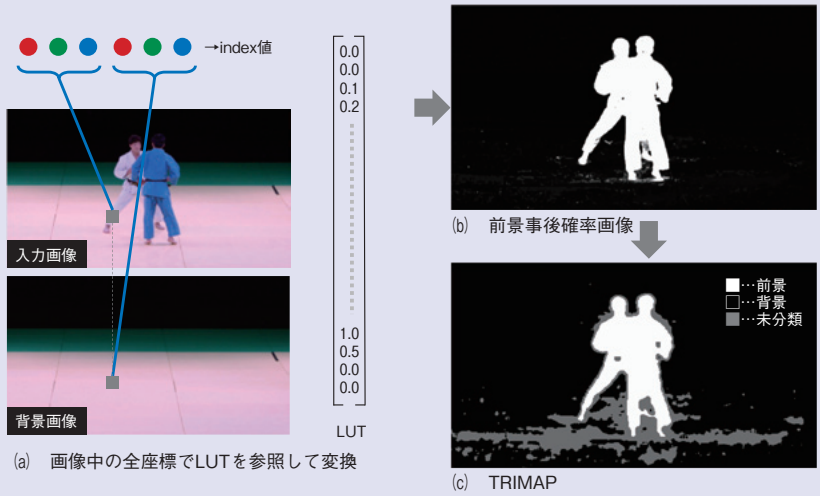
機械学習処理は事前学習処理と適用処理に大別できます．事前学習処理は，教師データからNNモデル中のパラメータを学習させることを目的として実行されます．事前学習処理フローを**図2**に示します．事前学習処理では，まず，教師データを用意します．被写体の含まれない背景画像と被写体の含まれるサンプル画像をあらかじめ取得しておき，正解となるマスク画像を作成します．次に，作成したマスクの前景領域に対応するサンプル画像中の注目画素と，背景画像中の対応画素を組み合わせたものを入力特徴ベクトルとし，その組合せが前景領域であるとして学習させます．また同様に，マスクの背景領域に対応するサンプル画像中の注目画素と，背景画像中の対応画素の組合せについても，入力特徴ベクトルが背景領域であるとして学習させます．このようにして，ある入力注目画素に対応する背景画素の組合せに対して，それが前景領域であるか背景領域であるかを識別するNNモデルを得ます（図2 (a)）．しかし，NN処理は一般に演算量が多いことから，NN演算処理をLUT（Look Up Table: ルックアップテーブル）実装することで高速に処理します．そこで，入力特徴ベクトルを量子化処理によって少ない階調数に削減し，量子化された入力特徴ベクトルのすべてのNN出力組合せをLUTとして保持します（図2 (b)）．なお，図2では分かりやすいようにRGBの各画素について行う処理を記載しましたが，画像位置情報を入れるなど入力ベクトルは色情報に限定されません．

LUTを適用し，TRIMAPを生成する適用処理を**図3**に示します．適用は



(a) 前景領域・背景領域を識別するNNモデル　(b) NN出力組合せをLUTとして保持

図2　事前学習処理



(a) 画像中の全座標でLUTを参照して変換　(b) 前景事後確率画像　(c) TRIMAP

図3　粗いマスク画像生成処理

*1 TRIMAP: 既知の領域と未知の領域を示した領域マップ．既知の前景領域を白，既知の背景領域を黒，未知領域を灰色に設定します．

*2 マッティング処理：被写体を抽出するアルファマスクを導出する処理．アルファマスクは0から1までの値を持ち，入力画像に掛け合わせて抽出映像が得られます．

機械学習処理と同様の処理で量子化した入力特徴ベクトルを導出し，導出した特徴ベクトルに応じてLUT参照することで，注目画素の前景画素である事後確率を高速に導出します（図3(b)）．得られた前景事後確率画像から，前景か背景かがあいまいである領域に対しては未分類領域と設定することでTRIMAPを生成します（図3(c)）．TRIMAPの未分類領域に対しては，注目画素と特徴ベクトルの近い周辺の画素が，前景と背景どちらの領域と識別されているかの情報を用いて，未分類領域を識別する境界補正処理を実施します．境界補正処理の詳細について**図4**に示します．未分類領域の注目画素について距離が近い領域をらせん状に探索を行い，注目画素が前景と背景にどちらが似ているかでアルファ値を決定しています．導出されたアルファ値を使い，被写体は抽出されます．この境界値補正処理を導入することにより，画素情報だけでなく，周りの画素の情報を考慮した被写体抽出が実行できるほか，粗い切り出しは低解像度画像に対して処理を行い，境界補正のみすべての画素に対して実行するなど，処理の効率化も図れるフレームワークとなっています．

赤外線光を用いた同一色背景からのリアルタイム被写体抽出

機械学習を用いることで，NN内で自動的に高次の特徴空間に変換してくれるものの，色情報や形状情報を頼りに抽出するため，入力特徴ベクトルが同じ場合は理論的に分離不可能となります．そこで，新たな特徴を増やす試みとして肉眼で見えない赤外線を利用した可視光－赤外線カメラによる被写体抽出システムを開発しました．

可視光－赤外線カメラを用いた被写体抽出撮影環境を**図5**に示します．平行に設置された可視光カメラと赤外線

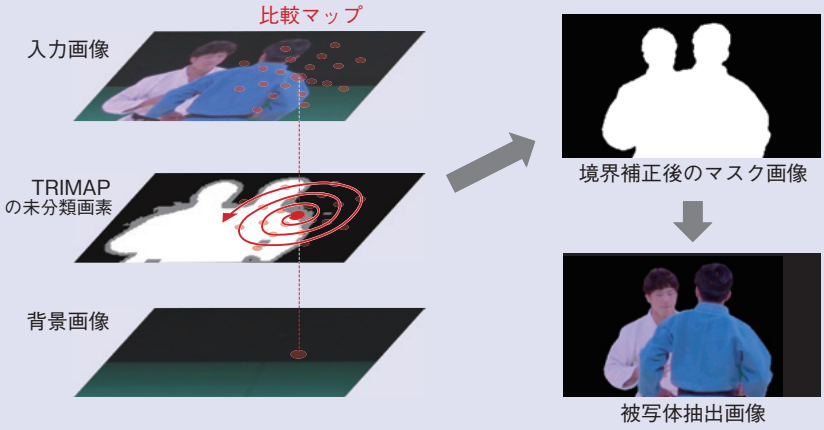


図4　境界補正処理とマスク合成処理

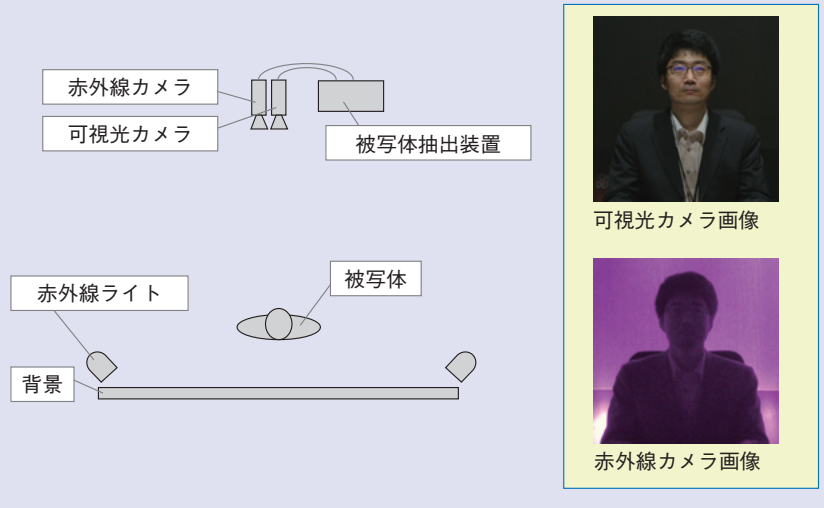
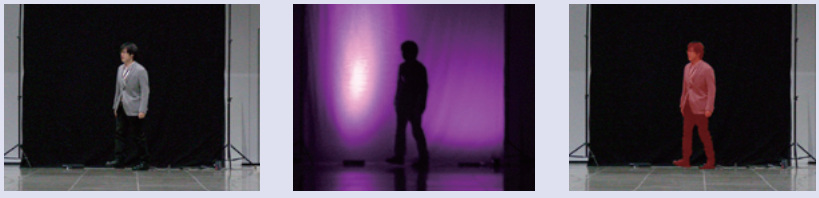


図5　可視光－赤外線カメラを用いた被写体抽出環境

カメラはそれぞれ，可視光のみおよび赤外線のみの画像を撮影します．赤外線ライトは背景に赤外線が照射され被写体には極力照射されないように設置します．これにより赤外線カメラ画像において背景は比較的明るく，被写体は比較的暗く写すことができます．図5の可視光カメラ画像では，人物と背景が同色のために色情報を用いた分離は難しいところ，赤外線カメラ画像では，人物のシルエットが抽出できていることが確認できます．次に，赤外線カメラ画像を精度良く被写体抽出するため，可視光カメラと赤外線カメラの視差補正を行います．補正は，あらかじめキャリブレーションボードを撮影しておき，赤外線カメラ画像と可視光カメラ画像の同一特徴点が重ね合わせられるように射影変換行列を導出・適用することで補正を行います．こうして得られた赤外線カメラ画像を機械学習の入力特徴ベクトルに入れることで，色情報や形状情報ではない新たな特徴を取り込んだ被写体抽出ができます．

(a)　入力画像　(b)　背景差分法による被写体抽出画像　(c)　機械学習による被写体抽出画像

図6　機械学習を用いたリアルタイム被写体抽出結果



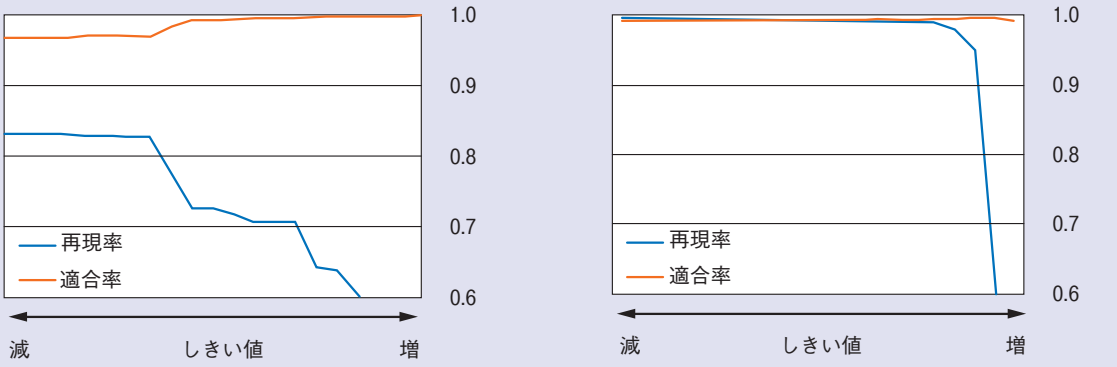
(a)　可視光カメラ画像　(b)　赤外線カメラ画像　(c)　被写体抽出結果

図7　可視光－赤外線カメラを用いた機械学習による被写体抽出結果

評価実験

機械学習を用いたリアルタイム抽出システムを開発し，今年4月に幕張メッセにて行われたニコニコ超会議2018の超歌舞伎「積思花顔競（つもるおもいはなのかおみせ）」の中で使用しました．開発システムは3840×2160画素，フレームレート60 fpsの映像に適用することができますが，他のシステムとの連携のために，本トライアルの中では1920×1080画素，フレームレートは59.94 fpsの映像に適用しました．超歌舞伎本編のクライマックスのシーンにて，中村獅童さん演じる惟喬親王（これたかしんのう）を背景が変化する中，舞台映像からリアルタイムに抽出し，初音ミク演じる小野初音姫（おののはつねひめ）の対決を盛り上げることができました．そのときの映像サンプルを**図6**に示します．惟喬親王の背景にある戸板は人手で持っているため揺れ，背景差分法ではうまく抽出できていませんが（図6(b)），機械学習を用いることにより正確に抽出できている様子が確認できます（図6(c)）．

次に，赤外線カメラを用いる効果を確認しました（**図7**）．赤外線カメラの利用により，可視光カメラだけでは抽出が難しい場合も被写体を正確に抽



(a)　可視光カメラ画像のみを利用した場合　　(b)　可視光カメラ画像と赤外線カメラ画像を利用した場合

図 8　背景差分法による再現率と適合率の比較

出できていることが確認できます．機械学習では，NNの学習の際に自動的に高次特徴空間への変換が行われ，しきい値処理などは発生しませんが，今回，赤外線カメラを用いることでどの程度頑健になるのか，背景差分法のしきい値を変化させることによって評価しました（**図 8**）．赤外線カメラを用いることで再現率・適合率がともに高く，また，背景差分法のしきい値を大きく変動させた場合にも安定して動作することが確認できます．

今回は，赤外線カメラを用いた場合を紹介しましたが，私たちのシステムは，ステレオカメラやLiDARから生成される深度マップなど，条件に応じて被写体をうまく抽出できる特徴を入力することもできます．

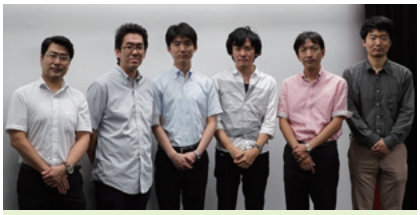
今後の展開

本稿では，機械学習を用い，入力特徴ベクトルを新たな特徴空間にNN内で変換・識別することで，被写体の高次の特徴量を基に高精度に抽出する方法，および可視光カメラだけではどうしても抽出することが難しいユースケースに対応するため，肉眼で見えない赤外線カメラを利用した手法について紹介しました．

今後は，セマンティクスを考慮した被写体抽出を行うため，深層学習を利用した被写体抽出のリアルタイム化を進めるとともに，被写体が重なるなどのオクルージョン発生時も精度良く被写体を抽出する方式の検討を進める予定です．

■参考文献

⑴　長田・宮下・柿沼・山口：“任意背景リアルタイム被写体抽出技術，” NTT技術ジャーナル，Vol.29，No.10，pp.33-37，2017.



（左から）長田　秀信/ 外村　喜秀/
宮下　広夢/ 柿沼　弘員/
日高　浩太/ 長尾　慈郎

見る人・使う人に驚きと感動をもたらすようなサービスやシステムの実現のため，今後も技術をさらに進化させていきます．被写体抽出技術を使うことで，今までになかったような観戦・観劇のスタイルや，SF映画で見たようなコミュニケーションスタイルが実現できると信じています．

◆問い合わせ先

NTTサービスエボリューション研究所
ナチュラルコミュニケーションプロジェクト
TEL　046-859-3780
E-mail　nagata.hidenobu@lab.ntt.co.jp