

VRMixer : 動画セグメンテーションによる動画コンテンツ と現実世界の融合

平井 辰典^{1,a)} 中村 聡史^{2,3} 湯村 翼³ 森島 繁生^{1,3}

概要: 本稿では、動画コンテンツと現実世界とを融合させることでユーザが動画の世界に入り込んだり、動画中の人物が現実世界に飛び出したりすることを可能とするシステム、VRMixer を提案する。本システムは、動画セグメンテーション手法によって動画の登場人物と背景とを別々に仮想空間内に配置することで簡易的な3次元空間(2.5次元空間)を構築する。構築された空間にデプスカメラによってキャプチャされたユーザを配置することによって、ユーザが動画の世界に入り込むことを可能とする。本研究では、動画から人物領域を自動で抽出するために、顔検出と人物領域検出を併用した3次元グラフカットセグメンテーション手法を提案する。抽出した人物領域を2.5次元空間の前方に、その他の背景領域を後方に配置することで仮想空間を構築し、デプスカメラによってキャプチャされたユーザは仮想空間内を自由に移動できる。本システムを通して、動画の中の世界と現実世界とが融合した世界、“Video Reality”の実現を目指す。

VRMixer: Mixing Video and Real World via Video Segmentation

HIRAI TATSUNORI^{1,a)} NAKAMURA SATOSHI^{2,3} YUMURA TSUBASA³ MORISHIMA SHIGEO^{1,3}

Abstract: This paper presents *VRMixer*, a system that mixes real world and a video clip letting a user enter the video clip and realize a virtual co-starring role with people appearing in the clip. Our system constructs a simple virtual space by allocating video frames and the people appearing in the clip within the user's 3D space. By measuring the user's 3D depth in real time, the time space of the video clip and the user's 3D space become mixed. VRMixer automatically extracts human images from a video clip by using a video segmentation technique based on 3D graph cut segmentation that employs face detection to detach the human area from the background. A virtual 3D space (i.e., 2.5D space) is constructed by positioning the background in the back and the people in the front. In the video clip, the user can stand in front of or behind the people by using a depth camera. Real objects that are closer than the distance of the clip's background will become part of the constructed virtual 3D space. This synthesis creates a new image in which the user appears to be a part of the video clip, or in which people in the clip appear to enter the real world. We aim to realize “video reality,” i.e., a mixture of reality and video clips using VRMixer.

1. はじめに

過去に撮影された動画や画像などのコンテンツは、記憶を振り返ることや、体験していない出来事の様子を知る上で有用である。古くはホームビデオやアルバムにまとめた写真を眺めることから、最近ではSNS上で画像や動画を

共有することなど、動画、画像コンテンツは長きに渡って人々の暮らしに密接に関わっている。一方で、これらのコンテンツを鑑賞する形態は大きく変わってはならず、すでに撮影されてしまったコンテンツは、ただ見ることでしか享受することができない。

日常生活では体験できない非日常を描写した映画などの動画コンテンツは長い間人々に親しまれてきた。しかし、映画などのスクリーンを通じたメディアでは、人々は観測者としてその世界を客観的に体験することしかできない。それでも人々の銀幕への憧れは強く、好きな映画のスク

¹ 早稲田大学
Waseda University

² 明治大学
Meiji University

³ JST CREST

^{a)} tatsunori.hirai@asagi.waseda.jp



図 1 VRMixer による動画コンテンツと現実世界の合成例

リーンの中へ飛び込みたいという思いを抱く人もいる。本稿では、既存の動画コンテンツを対象として、動画の世界と現実の世界を混ぜ合わせることで、ユーザが動画の中の世界に入り込むことや、動画中の人物を現実世界に飛び出させることを可能とするシステム、VRMixer を提案する。

映像の一部にユーザを登場させる代表的な技術として、クロマキー合成が挙げられる。クロマキーを用いた多くのシステムではユーザが動いている間も背景は静止している。このような背景と現実との間のギャップは、合成結果に対してリアリティを感じる上での障壁となる。クロマキーの背景を動画にすることで障壁は少なくなると考えられるが、動画内の時間はユーザの世界に流れる時間とは関係なく流れていき、依然としてギャップは残る。

VRMixer では、時間と空間の両方を共有することにより、ユーザがより深く動画の世界に入り込める空間を構築する。図 1 に VRMixer によって融合された動画と現実の合成結果のイメージを示す。例えばダンス動画内のダンサーと一緒にユーザがダンスすることで、ユーザはダンサーとの間に一体感を感じられると想定できる。動画内の世界と現実との間に明らかなギャップがあったとしても、ユーザが動画の一部となったり、動画内の人物が現実世界に飛び出したりすることによって、そのギャップは曖昧なものとなると考えられる。

動画の世界への没入感を実現するために、VRMixer は動画像処理によって動画コンテンツから 2.5 次元空間と呼ばれる簡易的な 3 次元空間を構築する。具体的には、顔検出と人物領域検出を併用した 3 次元グラフカットセグメンテーション手法によって動画内の人物領域を抽出する。これにより、動画中の人物領域と背景領域を自動で切り離すことが可能となる。2.5 次元空間は、人物領域を手前に、背景領域を奥に配置することによって構築される。デプスカ

メラで奥行きを計測することで、ユーザは 2.5 次元空間に入って、背景と人物の間や背景の後ろ、人物の前などを自由に動くことができる。動画の世界と現実世界の空間を共有することによって、ユーザの時間に対する意識は、動画の世界の中に流れる時間と混ざり、その境目は曖昧なものとなる。本システムを通して我々は、動画の中の世界と現実世界とが融合した世界、“Video Reality”の実現を目指す。

2. VRMixer の応用例

本章では、VRMixer によって実現可能なアプリケーションの例を挙げる。VRMixer は、仮想空間中の奥行き方向の情報を考慮することで、クロマキーを用いたシステムには実現できないような空間共有を可能とする。そのため、奥行き方向の情報を必要とするようなアプリケーションへの応用において特にその効果が発揮される。簡単な応用例の一つとして、ユーザが動画中の人物と仮想的な共演を果たせるようなアプリケーションが挙げられる。クロマキーによるコンテンツの場合には、ユーザは常に登場人物の前方に合成されてしまうが、VRMixer ではユーザが登場人物の後方に立つことも可能とする。そのため、ユーザはコンテンツの中で、時に脇役として参加することも可能となり、より平等な共演を実現する。VRMixer を使って映画やドラマ等のコンテンツへ入り込むことで、演技や芝居の練習をすることもできる。演技の練習は一人で行うことが困難であるが、VRMixer では既存のコンテンツに登場する俳優と仮想的に共演することで演技の練習が可能となる。

様々な隊形や動きの考慮が必要となるダンス動画への応用を考えたとき、奥行き情報の考慮は不可欠である。ニコニコ動画のような動画共有サイト上では、「踊ってみた」と呼ばれるユーザ（踊り手）によるダンス動画がアマチュアのダンサーの間で人気を博している。「踊ってみた」動画では、踊り手が音楽に合わせて踊った様子が動画共有サイトにアップロードされる。VRMixer を使うことで、ユーザは他の「踊ってみた」動画とコラボレーションしたり、好きなダンサーや歌手と一緒にダンスをすることが可能である。多くの「踊ってみた」動画では、踊り手はスタジオ等ではなく自身の部屋等のあまり魅力的ではない環境で動画を撮影していることが多い。VRMixer を使えば背景を他の動画で代用することでそのような問題も解決可能である。また、ユーザが実際のダンス動画の中に入ってダンスをすることはダンスの練習として有用であると考えられる。

3. 関連研究

動画の世界に入り込む手段として、Morishima らは Dive into the Movie という、ユーザを映像中に登場させることができる参加型映像コンテンツプロジェクトを提案した [1]。このプロジェクトの中で、Maejima らはユーザの簡易的な 3DCG キャラクタを作成して映像内に登場させる Future

Cast System を提案した [2]. Dive into the Movie は、映像コンテンツの新たな楽しみ方を提案したが、このシステムが適用可能なコンテンツはその専用のもに限られており、特別な装置なしには体験することができない。

コンテンツの世界を手軽に楽しむためには、一般的に手に入りやすいような最小限の機材でのシステム構築が望まれる。Hunter らは、遠隔地にいるユーザ同士の 3 次元空間内でのインタラクションを可能とするテレプレゼンスシステム、WaaZam¹を提案した [3]. WaaZam では、web 上に存在する 3D モデルや画像等のデジタルコンテンツを 3 次元空間内に配置し、その中にデプスカメラでキャプチャされた複数のユーザを仮想的に配置させることで新たな空間を構築する。Kinect の普及などにより、デプスカメラは一般的に手に入りやすいものとなっており、このような仮想空間内での体験はよりいっそう身近なものとなっている。WaaZam はテレプレゼンスシステムであり、ユーザが動画コンテンツ内に入り込むものではなく、遠隔地にいるユーザ同士のインタラクションを実現したものである。そのため、ダンス動画などのように登場人物の奥行きを考慮すべきコンテンツには適していない。また、WaaZam ではコンテンツの配置を手動で行わなければならないが、VRMixer では対象を人物が登場する動画に絞ることで、自動で仮想空間を構築することを可能としている。

VRMixer では、RGB ピクセル値にデプスチャンネルを加えた 2.5 次元空間により仮想空間を記述する。Kurihara らが提案した DDMixer2.5D では、2.5 次元動画を用いることにより、動画内のオブジェクトをドラッグ&ドロップして合成することを実現している [4]. 2.5 次元動画では、デプス情報があるため、オブジェクトのセグメンテーションが通常の動画に比べて容易にできる。2.5 次元動画お利用することで仮想空間の構築はより簡便化できるが、一般的に広まっている動画は多くの場合デプス情報を持っていないため、既存コンテンツの使用という観点からは 2.5 次元動画を用いた手法は適していない。そこで、我々は世の中に多く広まっている一般的な 2 次元 RGB 動画を対象とする。Li らは、3 次元グラフカットセグメンテーションを用いた 2 次元動画オブジェクトのカット&ペースト手法を提案している [5]. この手法は一般的な動画に対して適用可能であるが、我々の目的は動画同士の合成ではなく、動画と現実世界の合成である。2 次元動画から 2.5 次元または 3 次元動画を自動生成することは、現在の動画処理技術においても難しい課題である。Chen らは、2.5D paper cutouts を用いて動画を 2 次元から 3 次元へと変換する手法を提案した [6]. この手法の生成結果は高品質であるが、ユーザによる境界の指定やオブジェクトのデプスの指定など、多くの手作業が要求される。Iizuka らは、一枚の画像から 3 次元空間を復元する手法を提案したが、こちらもユーザによる手作業が要求される [7]. そこで、動画から

2.5 次元空間を自動的に生成するために、我々は主なオブジェクトを前方に、それ以外をすべて背景とする簡易的な 2.5 次元空間を構築する。本研究では、主なオブジェクトを人物と定義して動画内の人物の自動抽出を行う。この制約により扱うことのできる動画の種類は人物が登場している動画のみに限られてしまうが、その数は依然として多い。この制約に最適な、人物が常に主な役割を担っている動画として本稿ではダンス動画を主な対象とする。

4. システム構成

VRMixer は、2.5 次元空間を構築するために前処理を行う。図 2 に前処理を含めたシステムの全体像を示す。前処理は人手を一切介さずに行われる。手作業により動画内のオブジェクトを抽出するようなツールは多く存在するが、我々はこれらの作業をすべて自動化することで、システム導入の敷居を下げる。前処理が終わると、ユーザは動画内の世界に入り込むことができる。現実世界でデプスカメラの前で動くことで、動画の世界の中を動き回ることができる。合成結果はリアルタイムでディスプレイやプロジェクタなどの出力デバイスで表示される。このとき遅延はほとんど感じられないほどのもので、デバイスが画像を取得するための最低限な遅延のみである。

前処理 (動画セグメンテーションによる人物領域抽出)

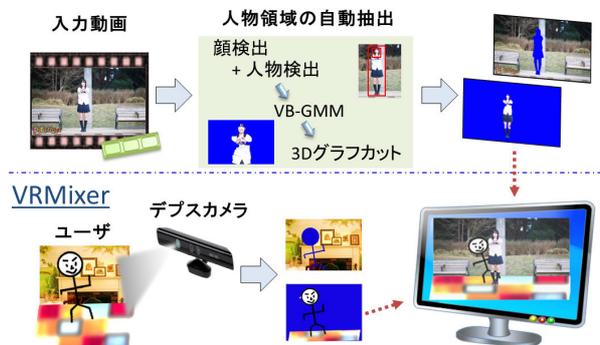


図 2 VRMixer の全体像

5. 2.5 次元空間の構築

5.1 動画セグメンテーション

本節では、VRMixer の前処理において最も重要な役割を担う動画のセグメンテーションについて述べる。VRMixer の前処理では、動画の各フレームは人物領域と非人物 (背景) 領域に分けられる。

5.1.1 人物領域の自動検出

動画中の人物領域の検出手法については、長らく研究されてきているが、学習なしでそれを実現することは難しい。HOG 特徴量を用いた手法 [8] が代表的であるが、特定の学習データが必要となり、対象の適用範囲が限られてしまう。

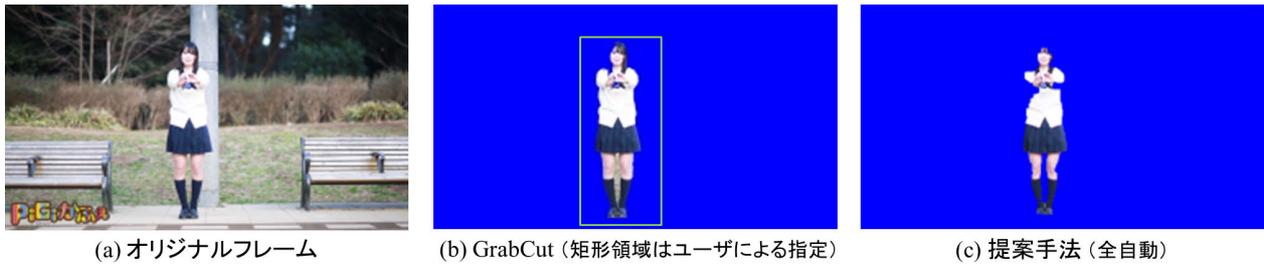


図 3 人物領域抽出手法の比較.

そのため、あらゆる服装や姿勢が考えられるような既存動画コンテンツでは適用は難しい。そこで、より幅広い既存コンテンツを対象とできるように、我々は顔検出手法を主な手法として使用し、人物検出手法は補助的に用いる。顔検出は、対象が未知の場合にも比較的高い精度で検出を行うことができる。顔検出手法には Irie らの ASAM に基づく手法を用いた [9]。顔検出は各動画フレームに対して行われ、顔の検出と同時に顔特徴点も検出される。顔特徴点により、大まかな顔の大きさ W を求めることで、顔から下 $n \times W$ の領域を大まかな体領域と推定する（現実装では $n = 8$ ）。同様に、顔から十分離れた箇所は背景領域であると推定できる。さらに体領域をより正しく推定するために、HOG 特徴量を SVM で学習させた人物検出手法を補助的に導入する [10]。これにより、顔検出には適していない横顔や後ろ向きの姿勢に対しても、人物領域を検出できる可能性が出てくる。このようにして推定した大まかな体領域と背景領域を基にして、より正解値に近い人物領域の推定を行う。

まず、大まかに検出した領域のピクセル値に対して VB-GMM [11] を適用して色分布を学習する。この色分布は、人物領域及び背景領域を構成する色の組み合わせとその強度を表している。これは、各動画に特化して人物領域及び背景領域を学習することに相当する。このようにして取得した色分布と大まかに推定した体/背景領域をシードとして min-cut/max-flow アルゴリズム [12] を用いた 3 次元グラフカットセグメンテーションを行う。3 次元グラフカットにより自然なセグメント境界を得ることができる。

さらに人物領域抽出の精度を上げるために、グラフカットを行う際に、フレーム内の動きの大きさに基づく重み付けを行う。これにより、ダンス動画などのような人物が動いている動画において、より正確に人物領域を検出しやすくなる。フレーム内の動きの検出は Gunnar Farneback アルゴリズム [13] を使った optical flow によって定量的に算出される。これらの処理に加えて、セグメンテーション領域の膨張・伸縮処理で細かいノイズを除去する。

5.1.2 動画セグメンテーション結果

図 3 に本手法による人物領域の自動セグメンテーション結果と、ユーザによる入力を用いたセグメンテーション手法である GrabCut との比較を示す [14]。このように、ユー

手法	適合率	再現率	F 値
GrabCut	0.719	0.856	0.768
提案手法	0.751	0.809	0.734

ザによる入力を用いた手法と比べても遜色のない結果となっている。さらに、人物領域精度を定量的に示すために、人物領域の抽出実験を行った。実験は、3 つのダンス動画からそれぞれランダムに 30 枚のフレームを選んできて、画像領域指定ツールを使って人手で作成した正解領域と、各領域検出手法との間のオーバーラップを評価することにより行う。表 1 に全 90 フレームの検出精度の平均値を示す。

GrabCut では、フレーム毎にユーザが人物領域を矩形で指定する。そのため、入力された矩形外は必ず背景領域として推定され、フレーム内のすべてを推定対象としている提案手法に比べて有利な条件である。これに加え、検出精度に大きな差がないこと及び提案手法が全自動である点を考慮すると、提案手法によるセグメンテーションは有効であると言える。しかしながら、検出結果には依然として誤検出が目立つことがある。これについては今後の課題として精度の向上を目指していきたい。

また、現在の実装では人物領域抽出の際に多くのパラメータ設定が必要である。例えば、動いている領域に加える重みの大きさは、人物が大きく動くダンス動画には有効であるが、人物がほとんど動かない動画では逆の効果を及ぼす。このようなパラメータのチューニングの代わりに、簡単なユーザの入力を受け入れることも一つの方法である。フレーム毎にオブジェクトが動く動画コンテンツでは、ユーザの入力は手間がかかってしまうことが多いが、Wang らが提案した interactive video cutout [15] や、Li らが提案した効率的な境界修正手法 [5] などの効率的なユーザの入力は検討の価値がある。

5.2 2.5 次元空間の構築

抽出した動画内の人物領域に関する情報を用いて、人物領域が前に、背景領域が後ろにくるように 3 次元空間内の異なるデプスに配置する。これに加えてデプスカメラでリアルタイムに抽出したユーザの情報を混ぜ合わせることで新たな 2.5 次元空間が構築される。VRMixer による動画

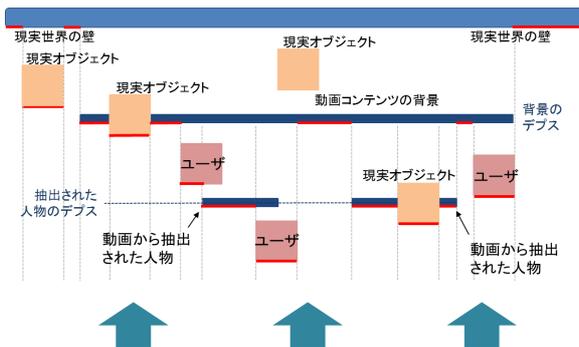


図 4 動画と現実世界との融合の模式図



図 5 VRMixer による合成結果

空間と現実空間の合成の様子を上から見た模式図を図 4 に示す。図中で赤線が引かれている箇所が実際に合成されるオブジェクトである。現実世界から合成されるものはユーザの身体に限らず、背景領域よりも前に配置されているものであれば合成結果に反映される。これにより既存動画のセットにユーザの好きなセットを加えることなどでもできる。

図 5 に VRMixer によって合成されたユーザと動画内のダンサーとのコラボレーションの様子を示す。ユーザは動画内の人物の前や、人物と動画の背景の間などに自由に立つことができる。現在の実装では、動画内の人物の奥行きや大きさはあらかじめ設定しなければならないが、今後、ユーザの奥行きと大きさに合わせて自動的に調整がなされるような現実世界と動画世界の正規化手法を導入したい。

6. VRMixer による動画と現実世界の融合

VRMixer によって、人物が登場する動画であればユーザが入り込むことが可能となった。しかし、VRMixer で実現可能な動画と現実世界の融合はそれだけではない。図 5 に示した合成結果は、実際にはユーザが動画の世界に入り込んだだけのものではない。合成された仮想空間において床は現実世界のものとなっており、現実世界の床に動画内の人物の足が染み出してきている。背景は動画の者であるが地面は現実のものとなっており、融合された空間において動画と現実の境界は曖昧なものとなっている。



図 6 動画内の人物をユーザの空間に呼び寄せる

6.1 動画内の人物を現実世界に合成

VRMixer では動画内の人物のみを現実世界に合成させることもできる。図 6 にその一例を示す。このような合成は AR 技術を使ったものでは一般的だが、VRMixer では特別な 3DCG モデル等を用意することなく、既存の動画をそのまま使用できる。この方法でユーザは好きな人物を好きな場所に呼び寄せることが可能となる。

6.2 ダンスの練習への応用

VRMixer の実用的なアプリケーション例としてダンスの練習への応用が挙げられる。ダンスの練習の際に、ダンス動画を見ながら振付を覚えていく人は多い。動画を見る際に、左右が反転しているため振付の向きを混乱してしまうことがある。VRMixer を使えば、実際に同じ空間でダンスをしている様子を見ることができ、左右の混乱は起こりづらい。さらに、ダンスの練習に特化させるために、鏡のように左右を反転させて表示する機能も実装されている。これにより、鏡の前で講師と一緒にダンスの練習をしているのと同様なことが、既存コンテンツ内の有名ダンサーとの間で実現できる。実際に遅延もほとんど感じられないため、VRMixer の合成結果は、任意のダンサーと一緒に踊れる魔法の鏡のような役割をも担える。フォーメーションダンスなどにおいても、奥行きが考慮されている分、ただ動画を見る場合に比べてより複雑な練習ができる。

7. 結果

VRMixer のユーザからのフィードバックを紹介する。あるユーザは、動画内の人物の足が現実世界の床に踏み入れることで動画と現実世界が混ざっていると感じたと述べている。またあるユーザは、動画セグメンテーションとデプスカメラの精度不足が目についたとコメントした。また別の意見として、ダンスグループの中のメンバーの一人に成り代わることができれば面白いという意見が得られた。これを実現するためには、動画内の人物を除去する手法を導入する必要がある。Kurihara らは PatchMatch アルゴリズムを使ったインペインティング手法 [4] を提案しており、VRMixer にも同様の手法が導入可能であり、今後

の実装で実現を目指したい。また、本システムそのものの定量的な評価については今後行っていく予定である。

7.1 今後の課題

動画セグメンテーションの精度は VRMixer にとって重要である。本稿で提案した動画セグメンテーション手法は顔認識に大きく依存しているため、顔検出の適用が厳しいような条件のフレームについては、フレーム全体が背景であると推定されてしまう。HOG を用いた人物検出手法により顔が検出できない箇所でも人物領域を検出できる可能性もあるが、それも万能ではない。今後、さらなる精度向上のために、ユーザによる簡単な入力とそれを効率的に反映するような手法についても検討していきたい。

シーンチェンジが多くあるような動画では、ユーザが入り込んだ世界が頻繁に変わってしまい没入感が得られづらい。また、全身が映っていないアップの動画の場合、現実世界との明確になってしまう。そのため、ダンス動画のようにシーンチェンジが少なく全身が映っていることが多いような動画が対象として適している。

将来的には二つ以上の動画との間の融合も検討している。それによりユーザがキャスティング可能な夢のコラボレーション等も実現できる。また、WaaZam[3] のようなテレプレゼンス機能を導入することで複数のユーザの環境とコンテンツとを混ぜ合わせることも可能となる。

また、ユーザと動画内の人物とのインタラクションの実現も検討している。例えばユーザが押す動作に合わせて動画内の人物が動けば、動画内の世界と現実世界との境界はさらに曖昧なものとなりうる。さらに、Tsuchida らが提案した自走式スクリーンなどを使えば、動画内の人物を物理的に呼び寄せることもできる [16]。

8. まとめ

本稿では、動画セグメンテーションを用いて現実世界と動画の世界とを融合させるシステム、VRMixer を提案した。VRMixer はユーザを動画の世界に入り込ませるだけでなく、動画内の人物を現実世界に呼び寄せたり、現実と動画の世界の境界を曖昧なものにさせることができる。このような動画と現実世界の融合は、動画コンテンツの新たな楽しみ方となりえる。過去のホームビデオを見て思い出を振り返る際などに、現在の自分が入ることで思い出を追体験するようなこともできるかもしれない。

インタラクティブデバイスの発展と共に、VRMixer によって提供可能な没入感も向上させられる。例えば、AIREAL[17] のような空気によるフィードバックが可能なデバイスを導入することで、動画内の人物の動きに合わせた風を感じさせることができ、それにより没入感もいっそう深まりうる。

我々は、VRMixer をあらゆる動画コンテンツに対応可能

なシステムとすることを目指している。過去の歴史的シーンの動画に入り込むことを可能にすることで、歴史を観測するだけでなく体験することができる。これはタイムマシンを仮想的に実現する一つの方法と言えるかもしれない。

動画コンテンツは重要なエンターテインメントソースであるにも関わらず、その消費の方法は昔からほとんど変わっていない。VRMixer の技術を発展させていくことによって新しいコンテンツとの関わり方の実現を目指したい。

謝辞 本研究の一部は、JST CREST「OngaCREST プロジェクト」及び、日本学術振興会特別研究員 (DC1) 科学研究費の支援を受けて実施された。

9. References

- [1] Morishima, S., Yagi, Y., and Nakamura, S.: Instant Movie Casting with Personality Dive into the Movie System, *Journal on Virtual and Mixed Reality*, pp.187–196 (2011).
- [2] Maejima, A., Wemler, S., Machida, T., Takebayashi, M., and Morishima, S.: Instant Casting Movie Theater: The Future Cast System, *IEICE Trans. on Information and Systems*, pp.1135–1148 (2008).
- [3] Hunter, S., Maes, P., Tang, A., Inkpen, K., and Hessey, S.: WaaZam!: supporting creative play at a distance in customized video environments, *Proc. of CHI2014*, pp.1197–1206 (2014).
- [4] Kurihara, T., Okabe, M., and Onai, R.: DDMixer2.5D: drag and drop to mix 2.5D video objects, *Proc. of UIST2013 Adjunct*, pp.69–70 (2013).
- [5] Li, Y., Sun, J., and Shum, H.: Video object cut and paste, *ACM Trans. on Graphics*, vol. 24, pp.595–600 (2005).
- [6] Chen, J., Paris, S., Wang, J., Matusik, W., Cohen, M., and Durand, F.: The video mesh: A data structure for image-based three-dimensional video editing, *Proc. of ICCP2011*, pp.1–8 (2011).
- [7] Iizuka, S., Kanamori, Y., Mitani, J., and Furui, Y.: Efficiently Modeling 3D Scenes from a Single Image, *IEEE Computer Graphics and Applications*, pp.18–25 (2012).
- [8] Wang, X., Han, T., and Yan, S.: An HOG-LBP human detector with partial occlusion handling, *Proc. of ICCV2009*, pp.32–39 (2009).
- [9] Irie, A., Takagiwa, M., Moriyama, K., and Yamashita, T.: Improvements to Facial Contour Detection by Hierarchical Fitting and Regression, *Proc. of ACPR2011*, pp.273–277 (2011).
- [10] Dalal, N., and Triggs, B.: Histograms of oriented gradients for human detection, *Proc. of CVPR2005*, pp.886–893 (2005).
- [11] Corduneanu, A., and Bishop, C.: Variational Bayesian Model Selection for Mixture Distributions, *Proc. of AI Statistics2001*, pp.27–34 (2001).
- [12] Boykov, Y., and Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, *IEEE Trans. on PAMI*, pp.1124–1137 (2004).
- [13] Farneback, G.: Improvements to Facial Contour Detection by Hierarchical Fitting and Regression, *Image Analysis*, pp.363–370 (2003).
- [14] Rother, C., Kolmogorov, V., and Blake, A.: “GrabCut”: interactive foreground extraction using iterated graph cuts, *ACM Trans. on Graphics*, vol.23, 3, pp.309–314 (2004).
- [15] Wang, J., Bhat, P., Colburn, R., Agrawala, M., and Cohen, M.: Interactive Video Cutout, *ACM Trans. on Graphics*, vol.32, 4, pp.585–594 (2005).
- [16] Tsuchida, S., Terada, T., and Tsukamoto, M.: A system for practicing formations in dance performance supported by self-propelled screen, *Proc. of AH2013*, pp.178–185 (2013).
- [17] Sodhi, R., Poupyrev, I., Glisson, M., and Israr, A.: AIREAL: Interactive Tactile Experiences in Free Air, *ACM Trans. on Graphics*, Vol.32, No.4 (2013).