

IFWB: 「第四の壁」を壊す簡易イメージベースモデリング手法

井阪 建 (非会員)

藤代 一成 (正会員)

慶應義塾大学大学院理工学研究科

IFWB: A Simple Image-Based Modeling Approach to Breaking the “Fourth Wall”

Takeru Isaka (Non-member) Issei Fujishiro (Member)

Keio University Graduate School of Science and Technology

isaka@fj.ics.keio.ac.jp

fuji@ics.keio.ac.jp



図 1: IFWB の概要

Figure 1: Overview of IFWB

概要

本論文では、ユーザが注目対象の画像 (図 1(a)) に対して臨場感をもち得ていない心理的状態「第四の壁」を、画像内の世界へのウォークスルー体験 (図 1(b)) を実現することによって壊すイメージベースモデリングシステムとして、IFWB (Instant Fourth Wall Breaker) を提案する。本システムは、奥行きのある構図をもつ画像を 1 枚入力すると、それ以上のユーザ入力を必要とすることなく即座に動作する。実際のターゲット環境として、スマートフォンやタブレット端末などの軽薄短小なデバイスを想定しており、カメラと同じような感覚で、いつでもどこでも使うことができる (図 1(c))。

キーワード: 第四の壁, Tour into the Picture (TIP), 運動視差立体視

Abstract

This paper proposes IFWB (Instant Fourth Wall Breaker) as a novel image-based modeling system that can provide a user with a feeling of presence in the world of image he/she is seeing. The user is required to input only a single image with a sense of depth (Fig. 1(a)), and then IFWB augments its depth without any further input and allows him/her to walk through the world of image freely (Fig. 1(b)). We mainly target at light and compact devices such as smartphones and tablets in order for the user to use the system readily just like a camera anytime and anywhere (Fig. 1(c)).

Keywords: Fourth wall, Tour into the Picture (TIP), motion parallax stereovision

本論文はNICOGRAPH2014における発表[1]の内容を拡張させたものである。

1. 背景と目的

人は絵や写真などを観賞する際、自分と対象画像との間に互いを隔てる不可侵な境界、演劇でいえば「第四の壁」に相当する限界をどうしても感じてしまうものである。絵や写真中の世界は本来自分とは無関係だと、鑑賞者は暗黙のうちに受け入れてしまっている。Bell (2008) [2]は、画像において「第四の壁」を壊すとは、「登場人物が視聴者に対して直接的に接触すること、または登場人物が視聴者に見られていることを自覚すること」と述べている。例えば、見ている画像中の人物が自分に向かって手を伸ばし触れてきたことを想定してみる。鑑賞者ははじめ驚くだろうが、その画像に対するそれまでの第三人称的意識とは違った、第一人称的意識を伴って応対すると考えられる。

この「第四の壁」という言葉に関して、これまで示されてきた意味や使われ方を総合的に考慮し、本論文では「観察者が、注目対象によって展開される世界の外部にいて感じてしまう心理的状态」と定義する。これを踏まえて、注目対象の世界内にあたかも自分がいるかのような感覚、すなわち注目対象を含む世界に対する臨場感を観察者に生じさせることによって「第四の壁」を打破することを考える。山本ら (2014) [3]は、テレビ画面に表示された映像に同期させたCGアニメーションを携帯端末に重畳表示することで、映像コンテンツがテレビ画面を越えて飛び出すような映像表現を可能にした。そこでは「第四の壁」という言葉は、人間の心理状態ではなく、テレビ画面という物理的存在をさしている。我々はこの研究のように画像コンテンツがこちらに飛び出してくるのではなく、情報を知覚する側であるユーザ自身が画像内に入っていくという体験を実現することで、ユーザが受ける臨場感を高め、汎用デバイス上で手軽に単一画像に対する「第四の壁」の打破を試みる。

安藤ら (2010) [4]によれば、臨場感とは単一の感覚ではなく複数の感覚要素から構成されており、空間要素、時間要素、身体要素の3つに分解できる。画像において、臨場感における空間要素を満たす効果的な手段として立体視がある。視差を利用した原理に関して立体視を分類すると、両眼の網膜上にできる物体像の差を利用した両眼視差立体視と、観察者の能動的な移動で生じる2時刻の網膜像間における変化量を利用した運動視差立体視に大きく分けられる。本論文では、後者を主要因としてユーザの奥行き知覚を刺激する。このとき、ユーザ頭部の移動および視覚刺激によって生じる自己運動感覚によって、臨場感における身体要素を満たす。さらに3.3節で詳述するが、臨場感における時間要素としてリアルタイム性を追求する。このようにして、「第四の壁」を打ち破るような臨場感のある画像を、入力された単一画像から半自動的にその場で生成するシステムIFWB (Instant Fourth Wall Breaker) を提案する。

被視対象の3次元形状と奥行きに関する情報は、運動視差のみで十分得られることが知られている[5]。運動視差立体視はさらに、メガネやヘッドマウントディスプレイといった装着デバイスを必要とせず、両眼視差立体視に比べて明度と解像度が落ちない自然な立体視を、ラップトップPCやタブレット端末

などの汎用デバイス上で手軽に実現できる利点をもつ。本システムはスマートフォン(以降、スマホと略す)に代表される軽薄短小なデバイス上で、ユーザがその場で取得した画像に対して気軽に利用することを狙っているため、その点では運動視差立体視を採用することは理に適っているといえる。ただし、それは表示するコンテンツがあらかじめ用意されている場合に限られる。

これまでの運動視差立体視を利用したシステム[6][7]は、表示対象をあらかじめ3次元オブジェクトとしてモデリングしておく必要があった。そのような環境の準備には当然手間や時間がかかるので、ユーザがその場で撮影した画像などには即座に適用できない。そこで、単一画像から視点を変更した画像を生成する手法であるTour into the Picture (TIP) [8]における透視制御パラメータを、汎用のウェブカメラによるユーザの顔追跡によってリアルタイムに再設定することで、運動視差立体視可能な画像を生成する。これにより我々の目標である即座かつ気軽に「第四の壁」を壊すようなユーザ体験を実現する。ユーザが眺めのよい場所にきたとき、即座に遠隔地にいる友人に単なる写真や動画以上に臨場感をもってその景観を伝えたい場合などにも、本システムは威力を発揮すると考えられる。

先行発表[1]では、入力として1点透視の構図をもつ画像のみしか扱えなかったが、本論文では2点透視の場合でも対応できるように拡張した手法を提案する。また、入力画像が3点透視図の場合も考察し、より広範な種類の画像に対応できる可能性を示す。さらに、各種の定量的なユーザ評価を行い、本システムの有効性を検証する。

2. 関連研究

本章では、5つの関連研究を採り上げ、本研究の意義と立ち位置を明確にする。

本システムが手軽さと即時性を獲得するために最大限に利用しているのがHorryら(1997) [8]のTIPである。イメージベースレンダリングが全盛で、複数画像を用いて3次元形状を復元することが主流だった当時、1枚の入力画像のみで違和感のない自然な3次元復元を達成した。TIPではスパイダリメッシュの位置および大きさの指定が必要であるが、IFWBではそれを半自動化し、より一層ユーザの負担を軽減している。

より詳細な3次元構成という面において、TIPの拡張手法が数多く発表されている。そのなかで代表的なものがLizukaら(2012) [9]の手法である。ユーザに床面と壁面の境界を入力させることでTIPよりも複雑な3次元構成が可能であるが、ユーザ入力の自動化が困難で、簡便さという面は考慮されていない。また、複雑な3次元構造をもたせた場合に、インタラクション方法も同様に複雑になってしまうという懸念も残る。

TIPの拡張に関して、簡便さを追求した手法としてBoulangerら(2006) [10]のATIPがある。ユーザ入力をTIPにおけるスパイダリメッシュの大きさ調整のみに留めていながら、多点透視図、消失点が画像内に存在していない場合など多種類の入力画像に対応できる。IFWBでは、対応できる画像は限ら

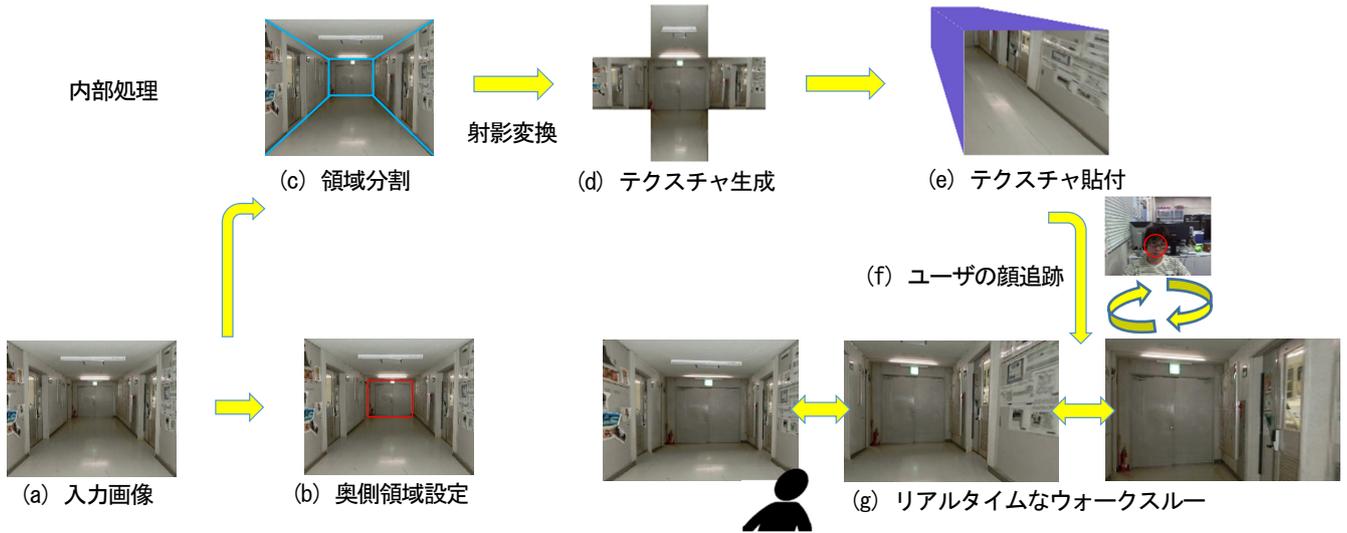


図2: IFWBにおける処理の流れ

れるものの、ATIP よりもさらに簡便さを追求している。運動視差立体視を利用してユーザに立体感を与えるシステムとして、代表的なものに末永ら (2005) [6]の研究がある。このシステムでは、あらかじめ3次元空間上でモデリングされたオブジェクトが配置されたバーチャル空間の前にユーザがいるという設定のもと、1組のステレオカメラを用いて顔追跡を行う。計測されたユーザ頭部の3次元位置、姿勢、視線ベクトルによりCG画像を透視投影変換することで運動視差立体視を可能にしている。しかし、コンテンツのモデリングが必要という点と顔追跡専用デバイスを用いているという点で、即時性と手軽さは失われている。また、末永らはこの研究を拡張させた研究[7]において、入力に多視点画像データベースを用いることで、モデリングされたオブジェクトを不要としているが、この場合も前もって被視対象の画像データベースを作成しておかなければならず、手間やコストがかかってしまう。

本システムはTIPにおける簡便さをさらに追求し、それを汎用のウェブカメラを用いた顔追跡を組み合わせることにより、軽薄短小なデバイス上でも手軽に画像内の世界にいるかのような、臨場感のあるユーザ体験を提供するものである。

3. 手法

図2にIFWBにおける処理の流れを示す。TIPでは、入力画像に対して透視図法における消失点と、ユーザの視点から見て奥の領域となる矩形を設定することで、入力画像(図2(a))を奥と上下左右の5つの領域(スパイダリメッシュ)に分割し、テクスチャを射影変換した後、再配置する(図2(b), (c), (d))。TIPではユーザが手動で消失点と視点を変更し、そのたびにテクスチャの射影変換を行っている。一方本手法では、図2(e)のように、3次元空間における直方体の奥と上下左右の壁に、スパイダリメッシュをテクスチャとして貼り付けることで、比較的重い射影変換を前処理に限定できる。そして3.3節で詳しく説明するユーザの顔追跡(図2(f))によって、ディスプレイに

対するユーザ頭部の相対位置に応じて、直方体と同空間上のバーチャルカメラ(以降、カメラと略す)を制御し、リアルタイムに消失点と奥の矩形の位置を設定している(図2(g))。能動的なユーザ頭部の動きに応じて画像が自動的に変化するので、ユーザは単一の画像に対しても、運動視差による立体感と自己運動感覚をリアルタイムに得ることができる。以上がIFWBの基本的な処理の枠組みであるが、さらにユーザの視覚に訴える強調表示機能を加えた。これについては3.4節で詳しく述べる。

3.1. 奥側領域の推定

入力画像として、前景物が目立たない1点透視図または2点透視図に近似できる画像(図3(a))を用いることを仮定する。高い頑健性をもつCanny法[11]によりエッジ抽出した入力画像(図3(b))に対して、図3(c)に示す矩形領域の形状および、矩形から延びる4線分の長さや角度を変化させたテンプレートを複数用いたテンプレートマッチングを行い、最も類似度が高い入力画像上の領域を検出することで、TIPではユーザ入力が必要であった奥側の矩形選択を半自動化した(図3(d))。

検出された最も高い類似度をもつ矩形領域に対して、あらかじめ設定した閾値以内の類似度をもつ矩形領域が他に検出されないとき、入力画像を1点透視図と判断する。閾値以内の類似

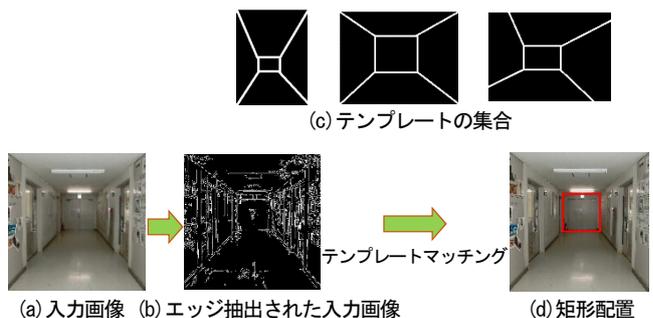


図3: 奥側領域設定の流れ

度をもつ矩形領域が存在し、かつ共有領域をもつ場合、それらの矩形領域群を奥側領域候補として提示し、ユーザの判断で最適なものを選択させる。

最も高い類似度に対して閾値以内の類似度をもち、かつ互いに共有の領域をもたない矩形領域が存在するとき、入力画像を2点透視図と判断し、次節で述べるように3次元構造を与える。

また、提案手法が手動でユーザが奥側領域を指定する場合と比べて、ユーザが受ける3次元構成の印象にどれほど差異が出るかを検証した。その詳細は5.2節で述べる。

3.2. 3次元構造の付与

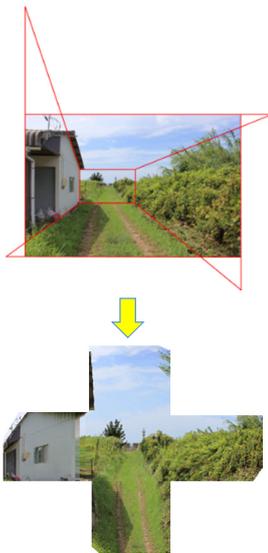


図4: 入力画像の4隅にスパイダリメッシュの境界線がこない場合の領域分割

図2(c)のように、前節で位置決定されたプレート矩形の4隅とそこから延びる線分の延長線と交わる画像の端を直線で結び、5つの領域に分割する。本手法では、図2(a)のようなスパイダリメッシュの境界線が画像の4隅にくるような入力画像が望ましい。図4のように、スパイダリメッシュにおける4つの線分の延長が入力画像4隅にはこない場合、それらの線分が画像端の延長線と交点をもつまで、線分を延長することで得られる矩形および台形形状で画像を領域分割し、スパイダリメッシュを初期化する。次に、奥側矩形の幅、入力画像の幅および一般的と想定されるユーザ頭部とディスプレイ間の距離を用いて、相似の原理から幅、高さおよび奥行きを決定した直方体を3次元空間上に生成する。

スパイダリメッシュの各領域を、この直方体のそれぞれ対応した面の形状に合うように射影変換したのち、テクスチャとして内部に貼り付ける(図2(e))。単純な3次元構成方法ではあるが、4章で示すように、一見透視図の構図をもたないような画像に対しても有効にはたらき、立体感が付加されている。

図5(a)のような構図をもつ画像を入力し、3.1節の奥側領域設定の処理において2点透視図と判断された場合、最類似プレート同士の矩形から延びたエッジの交点を結ぶ線分によっ

て、各消失点に属する領域が干渉し合わないよう入力画像を分割する(図5(b))。それぞれの領域に対して、前述したフローにしたがって直方体を2つ生成し、各領域のテクスチャをそれぞれマッピングする(図5(c))。また、この交わるエッジ間がなす角度によって、2直方体間が共有するエッジ(図5(b)の紫色の線分)を軸として、2つの消失点間の距離に応じた角度だけ左右の直方体をそれぞれ反時計および時計まわりに回転させる。



(a) 入力画像

(b) 奥側領域推定



(c) 直方体生成およびテクスチャ割当て

図5: 2点透視図法における3次元情報付与処理

3.3. 顔追跡による視点の変更およびウォークスルー

図6のように、前節でテクスチャが貼り付けられた直方体と同一空間上に置かれたカメラの位置と角度を、ユーザの顔追跡によって制御することで、リアルタイムに視点を変更している。図6右上にはユーザ頭部が表示されている。ユーザがディスプレイに顔を近づけたり、遠ざけたりすると、ユーザの顔の範囲を示す円の直径が変化する。この値を用いて、カメラの奥行き座標を制御し、ウォークスルーを実現している。ただし、前節での3次元構成の際に、直方体手前において意味あるピクセル値が入力されていない領域が存在する場合がある。ウォークスルー中にこのような領域が視野に入ると、ユーザの臨場感をそいでしまうおそれがある。そのため、カメラ画角内の全領域で意味あるピクセル値が入力されているカメラの奥行き座標を可動領域の始点とする。さらに、被視対象がカメラの視野から外れないように、ある程度カメラモーションに制限をかけている。

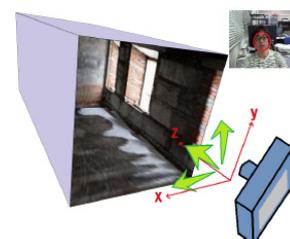


図6: 視点移動時のカメラ動作

本システムでは、正確でありながら高速に対象物体を検出できる Viola-Jones 法[12]を適用し、ユーザの顔追跡を行っている。図 2(f)のようにウェブカメラで得られた画像内からユーザの顔を検出すると、顔と推定される領域に円が表示される。この中心座標と直径によるフレーム値を用いることで、TIP 処理部におけるリアルタイムな視点変更の処理を制御する。そのままの値を用いるとカメラの挙動が断続的になってしまうので、得られた直近の 12 フレーム値をバッファリングし、時系列で直近の値ほど重みをつけて平均をとることで、ユーザの顔の位置座標を連続的に変化する 3 次元ベクトルとして扱う。

上記のままでも良好な顔追跡を行えるが、ウェブカメラで得られたフレームに対して逐次処理を施してから描画を行うと、どうしても実際の顔動作と生成画像間に遅延が生じてしまう。1 章で述べたように、臨場感における時間要素としてリアルタイム性があり、そのままでは臨場感が損なわれてしまう。さらに、実際の動作から遅れて画像が更新されると、VR 酔いを誘発しやすく、ユーザ体験の質を著しく下げてしまう。

そこで、既知のフレーム値から次の瞬間の顔位置を予測し、未来のフレームを補外することを考える。Lagrange 多項式を補間多項式として用いて、未来のフレームも含めた任意のフレームにおける座標値を曲線で表す。Lagrange 補間は入力点数が多くなると振動してしまうので、入力に用いるフレーム数を限定する。既知の 12 フレームを時系列順に 3 フレームずつ 4 つに分け、時系列順において直近のものほど値に重みをつけて平均をとり、代表値を求める。 α フレーム目の値に乘算する重み ω を式(1)に示す ($1 \leq \alpha \leq 12$)。式(2)において、 x_k は 3 つずつ分割された 4 つのフレーム値の集合内での、それぞれの x 座標値の重み付き平均値 ($0 \leq k \leq 3$) を表す。

$$\omega = 1 + \frac{\alpha}{12} \quad (1)$$

$$L_k(x) = \frac{x-x_0}{x_k-x_0} * \dots * \frac{x-x_{k-1}}{x_k-x_{k-1}} * \frac{x-x_{k+1}}{x_k-x_{k+1}} * \dots * \frac{x-x_3}{x_k-x_3} \quad (2)$$

式(2)で与えられる関数 $L_k(x)$ を定義したとき、13 フレーム目の x 座標 $L(x)$ は次の式で求まる。

$$L(x) = \frac{\sum_{k=0}^3 L_k(x) * x_k}{4} \quad (3)$$

式(3)で表される Lagrange 補間計算を x, y, z に関してそれぞれ行い、13 フレーム目にあたるフレーム値を算出した。

比較実験を行った結果、補外がある場合の方がユーザ操作の速度は向上した。詳しい結果は 5.3 節に示す。以上のようなソフトウェア上の工夫をすることにより、既存研究[6][7]のように専用デバイスを使わず、汎用のウェブカメラを用いて遅延の少ない画像更新を可能にした。

3.4. 前景物および余白による臨場感の強化

本システムで構成された 3 次元空間には前景物が存在していない(図 7(a))。既存研究[6][7]ではバーチャル空間上に置かれた前景物が立体視のうえで、有効な奥行き手がかりの一部にな

っている。そこで、本手法で構成された 3 次元空間上に図 7(b)に示すような 3D モデルを配置することで、ユーザに与える立体感を高めることを考える。

既存研究において、被視対象である部屋の中に前景オブジェクトが存在するような状態の場合、視点をずらした際、前景オブジェクトがディスプレイの端にかかって完全に表示されず、ユーザにディスプレイの存在を感じさせてしまい立体感を損なうことがある。そこで、図 7(c)のように画面全体を直方体の表示に使わず、画面の周囲にあえて余白を設けることで、本来なら前景オブジェクトが欠けて見える場合も、ある程度の視点移動に対しては、画面上に内部オブジェクトが完全に表示することが可能になる。ただし、余白を設けることで背景の表示領域を狭めてしまうことになる。4 章に示すようなラップトップ型 PC を用いてユーザ評価を行った際、『背景に対してキャラクタが飛び出して見えた』など立体感を増強しているとの感想をもらい、不満な意見は出なかった。さらに 5.4 節における定量評価でも臨場感を高めていることを確認できた。しかし、本システムが実行環境と想定している、スマホ等の表示領域が狭いデバイス上では、臨場感がかえって削られる可能性が残る。



(a) 前景物, 余白共になし



(b) 前景物あり, 余白なし



(c) 前景物, 余白共にあり

図 7: 前景物と余白の有無による実行結果の違い

4. 結果

開発環境として、PC (CPU: Intel Core i7-4770K 3.40GHz, GPU: NVIDIA GeForce GTX 650Ti, 実装メモリ: 8GB, ディスプレイサイズ: 23.8 インチ), ウェブカメラ: Logicool Webcam C910 (平均フレームレート: 約 30 frame/sec) を用い、プログラミング言語に Delphi, 顔追跡のための API として OpenCV2.3[13] を使用した。

本システムを起動すると、図 7(c)のように画像の下にボタンがいくつか配置された画面が立ち上がる。画像読み込みボタンをクリックすれば入力画像を選ぶことができる。望みの画像が表示されている状態で、展開ボタンをクリックすれば即座に、または奥側領域候補が表示されれば、そのうち 1 つを選択したのちウォークスルーが可能である。ユーザが手動で奥側の矩形領域を選択することもできる。その場合、選択したい領域の対角線をドラッグしたのち展開ボタンをクリックすればよい。

実際に体験しなければ本手法本来の効果は得られないが、図 8 に 3 例の入力画像に対する結果を画像で示す。上から 1 例目と 2 例目の左上部に表示されたユーザ頭部の位置に応じて、画像が更新されていることがわかる。最下行にはルシニョール(S.

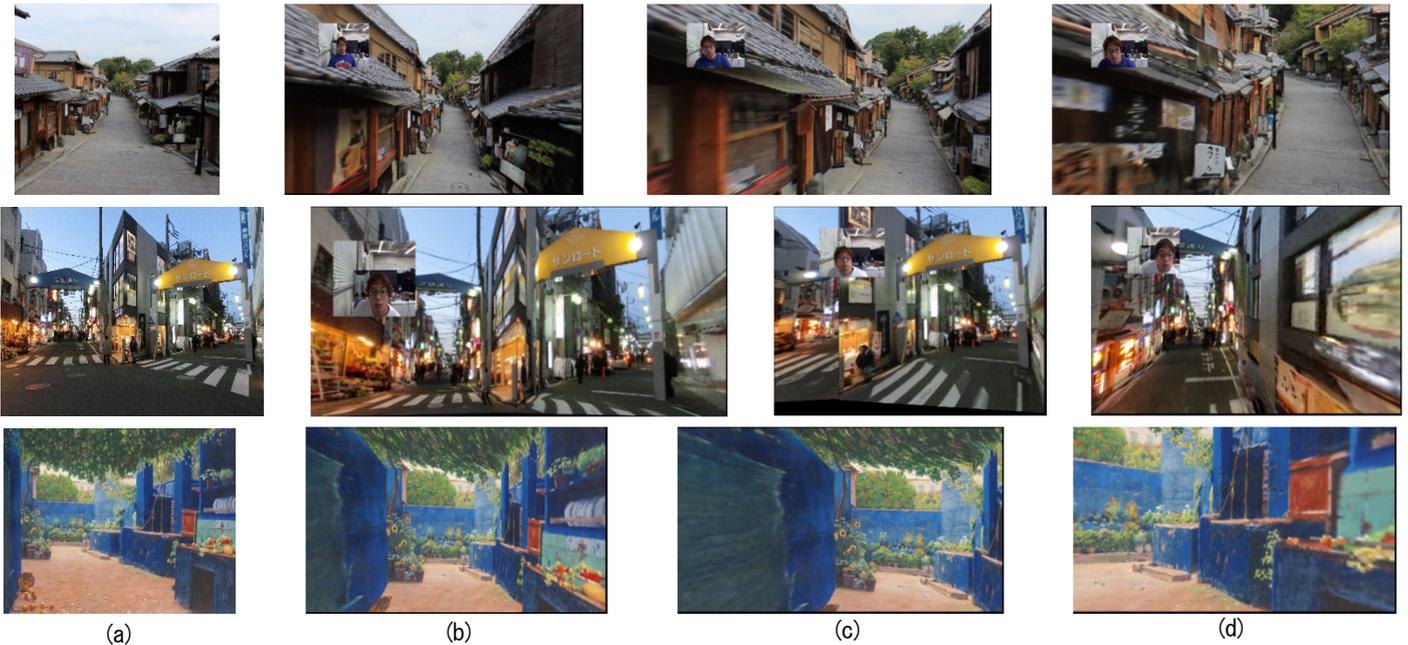


図8: 実行結果 — 写真2例と絵画1例(最下行)を入力画像としたときの実行結果を示す。最上行と最下行の例は1点透視図を, 中行の例は2点透視図を入力画像として使用した。それぞれ左から順に, 入力画像(a), ウォークスルーした際の遠景(b), 中景(c), 近景(d)である。最上行と中行の左上部にはユーザの顔位置を示している。中行の例では, 中景と近景でそれぞれ異なる消失点の方向にウォークスルーしている。

Rusinol) の《アレニス・デ・ムントの青い中庭》という絵画[14]を入力画像として使用した場合が示されている。この結果画像から, 実写画像だけにとどまらず, リアリティの薄い芸術作品においても立体感が得られていることがわかる。

5. 評価

提案システムが結果として意味のあるものかを検証するため, 前章で述べた実装環境のもとで, 3次元構成, 顔追跡, 余白の効果, 臨場感の4項目に関して評価を行った。以下, それぞれの詳細について述べる。

5.1. 定性的評価

本システムを2014年12月5日に行われた, 著者らが所属する慶應義塾大学理工学部が主催する技術展示会 KEIO TECNO-MALL2014 (<http://www.kll.keio.ac.jp/ktm/>) に出展し, 定性的なユーザ評価を行った。実際に本システムを約30人の来場者に試してもらったところ, 『一枚の写真から生成したとは思えない』, 『本当に絵の中に入っているようだ』などの驚きや肯定的な意見を25人からもらった。一方で, 3次元構造の正確性や詳細性に欠けるという意見を3人からもらった。直方体の手前側でテクスチャの粗さが目立っているとの指摘も受けた。またウォークスルーの際, ウェブカメラの画角を外れた頭部動作, ウェブカメラに顔が正対しないことで顔追跡が中断されてしまう例が見受けられた。ユーザ頭部を3Dモデルとして扱うなど, インタクション方法にも改善の余地があると考えられる。

5.2. 3次元構成の評価

本手法では, TIP では本来必要な入力画像に対するユーザ入力が必要としないため, 3次元構成の正確性に欠けている可能性が考えられる。そこで, 表1右に示す一般的な1点透視構図をもった画像9枚, 2点透視構図をもった画像2枚に対して, ユーザがスパイダリメッシュの奥側領域を手動で入力した場合と, 3章で述べた手法で半自動化した場合を比較して, ユーザが受ける3次元情報の印象に差異がどれほど表れるか検証を行った。実験参加者5人(いずれも正常な視覚をもつ20歳代男性4人および女性1人)に対して, それぞれの方法で生成された画像を20秒間提示した(図9)。「提案手法で生成された3次元空間に対して, 手動で入力した場合と比較して同じ印象を感じたか」(Q1とする), 「提案手法で生成された3次元空間に対して, 手動で入力した場合と比較してより適切な3次元構成が行われていると感じたか」(Q2とする)という2つの質問について5段階(感じない, どちらかというと感じない, どちらでもない, どちらかというと感じる, 感じる, の評語群を用い, それぞれ順に-2, -1, 0, 1, 2の点数を割り振った)で回答させた。

その結果を表1に示す。Q1に関して, 手動と半自動の場合で生成された3次元構成に対して異なる印象をもっている実験参



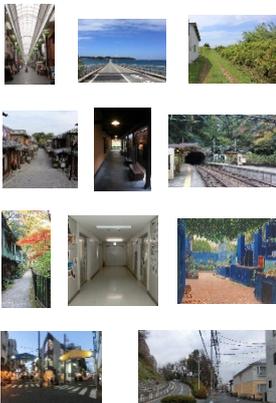
図9: 3次元構成に対するユーザ評価の様子

加者がやや多いが、Q2において手動の場合よりも提案手法による3次元構成のほうが適切であるという評価を得た。

ただし、これは実験参加者が本手法に関して既知かどうかの影響していると考えられる。奥側領域を画像面積に対して極めて小さい範囲で選択してしまうと、射影変換の際奥側領域以外のテクスチャが引き伸ばされてしまう。また画像の入力に対して一定の熟練度が認められ、後半になるほど選択領域が適切であるように感じた。したがって3次元構成の精度に関して、本システムについて既知のユーザや慣れたユーザは手動入力の方が良いが、初見のユーザに関しては半自動入力の方が適切だと考えられる。総合的に考えて、3次元構成の半自動化によって立体感の印象は損なわれていないと判断できる。

表1: 3次元構成のユーザ評価結果 (実験参加者数: 5) と対応する入力画像群、左上から順に画像1, 画像2, ..., 画像10, 画像11

	Q1	Q2
画像1	-1.2	1.2
画像2	-1.6	0.6
画像3	-1.4	1.75
画像4	-2	0.6
画像5	-1	-0.6
画像6	0.8	-0.6
画像7	2	-0.4
画像8	1.2	0
画像9	0.6	-0.4
画像10	-1	0.1
画像11	-1.2	0
平均得点	-0.43	0.20



5.3. 顔追跡の評価

3.3節で述べたように、顔追跡に関して、フレーム値の補外の有無でユーザ操作がどのように変化するか検証するため、比較実験を行った。図10のように底面にランダムな8桁の数字が書かれた直方体(赤色のオブジェクト)を3次元空間上に配置し、ユーザがその数字を覗きにいき、読み取れるまでの時間を計測するという実験である。

表2に実験参加者11人(全員正常な視覚をもつ20歳代男性)の平均時間を示す。補外ありの場合では、補外なしの場合と比較して約13%速いことがわかった。



図10: 比較実験用のバーチャル環境

表2: 補外の有無によるユーザ操作速度の比較 (実験参加者数: 11)

	補外なし	補外あり
平均時間 (秒)	15.84	11.15

5.4. 前景物および余白追加の評価

3.4節で述べた臨場感を高める目的で行った強調表示に関して、その効果を確かめるため次のような実験を行った。

表1の画像9を入力とし、8人の実験参加者(いずれも正常な視覚をもつ20歳代男性6人および女性2人)に前景物は追加した状況で、余白を追加しない画像と追加した画像をそれぞれ20秒間ずつ提示した。その後、「体験した画像に対して、条件がある場合とない場合でどちらがより画像内の世界にあたかも自分がいると感じたか」という質問に対して5.2節と同様な形式で回答させた。実験参加者には事前に条件を変更した旨を伝えたが、前景物および余白ということには言及しなかった。前景物は図7(b)に示すような3Dモデルを用い、カメラの始点座標からみて左手前に配置した。

結果を表3に示す。条件変更後ではユーザに与える臨場感が約17%増加することがわかった。ただし、画像と前景物の種類および配置位置などによる影響を今回は考慮していない。

表3: 前景物および余白追加によるユーザ評価の結果 (実験参加者数: 8)

	余白なし	余白あり
平均得点	1	1.17

5.5. 臨場感の評価

1章で述べたように、ユーザに臨場感を生じさせる有効な指標として自己運動感覚がある。瀬川ら(2003)[15]によれば、視野上のオプティカルフローや絵画の手がかりから人は視覚誘導性の自己運動感覚を感じる。これは周辺視野が特に重要な役割を果たしており、ラップトップ型PCや想定しているスマホなどにおけるサイズの小さいディスプレイでは、生じる効果が薄いと考えられる。提案手法では、顔追跡によるリアルタイムな描画更新によって実際にユーザ頭部の移動を促すため、視覚情報に加えて平衡感覚をつかさどる卵形囊や迷路小囊からの前庭系情報によって自己運動感覚が生じていることが期待できる。この仮定を検証するため、7人の男女(いずれも正常な視覚、前庭感覚を有し、年齢は20歳代男性5名、女性2名)を対象に次のような実験を行った。

提案手法によるシステムとカメラ操作をマウスで行わせるシステムを用意し、「画像内の世界を自由に観察してください」というタスクを実験参加者に与え、体験時におけるカメラの奥行き方向の速度を計測した(前進時を正の速度とする)。表1の画像8を入力として、画像提示は20秒間行った。マウスによるカメラ操作方法(表4)については事前に実験参加者に説明を行った。また画像自体の印象で操作に影響が出ることも考えられるため、表1の画像9に対しても同一の検証を行った。同時に「画像内の世界にあたかも自分がいると感じたか」(Q3とする)という質問に対して5.2節と同様な形式で回答させた。

表 4: マウスでのインタラクション方法

カメラの挙動	対応するマウス操作
ロール	画面をクリックしながら左右に動かす
ピッチ	画面をクリックしながら上下に動かす
前進 後退	それぞれホイールを前向き、 後ろ向きに回転させる

表 1 の画像 8 を用いた実験結果を図 11 に示す。各グラフの凡例は実験参加者を表す。例外も存在するが、マウス操作の場合 (図 11 (a)) では、静止している時間が長く、速度も加減速が急である。顔追跡を用いた操作の場合 (図 11 (b)) では、こちらも例外はあるものの、マウス操作の場合と比較してそれほどバラつきのない速度でウォークスルーを行っている様子が見て取れる。表 1 の画像 9 を入力とした場合でも似た傾向の結果が得られたため、ここでは省略する。加藤ら (2005) [16] は、自己運動感覚が有効な状況では、知覚される 3 次元情報の変化に応じてユーザは適切に加減速を行うと述べている。このことから図 11 (b) が示す比較的滑らかな速度変更は、顔追跡を用いたウォークスルー体験によってユーザに自己運動感覚が生じていることを示すと考えられる。また、Q3 に対する実験参加者 7 人の平均得点を表 5 に示す。マウスを用いたシステムではどちらの入力画像に対しても負の値を得る結果となったが、提案システムでは 1 前後の値を得られた。これらの結果から、提案手法では得られる自己運動感覚が一助となり、ユーザは臨場感を知覚していると判断できる。

表 5: マウス操作と提案手法による臨場感の評価結果 (実験参加者数: 7)

	マウス	提案手法
表 1 の画像 9	-0.69	0.86
表 1 の画像 10	-0.71	1.29

5.6. 考察

以上のような評価、意見から判断すると、単一の写真や絵画といった画像コンテンツに対する「第四の壁」を、その場で即座に取り除くという本研究の初期的目標は達成できたと考える。

今回、ディスプレイのサイズで、ユーザ体験にどのような影響が出るかは調査していない。しかし、ラップトップ PC 環境と比較して軽薄短小なデバイスを用いた場合では、ユーザ頭部はディスプレイに接近するので、視野における相対的なディスプレイ占有面積はそれほど変化せず、ユーザに生じる視覚誘導性自己運動感覚の知覚量も大きな差は出ないと予想している。

6. 結論および今後の課題

本論文では、TIP と汎用のウェブカメラによる顔追跡を組み合わせて、軽薄短小なデバイス上で、ただ 1 枚の画像を入力として手軽かつ即座に臨場感あるユーザ体験を実現する初期的システムを提案した。

以下では、本システムの機能を増強するうえで、次の段階で取込みを考慮すべきさまざまな拡張要素について考察する。

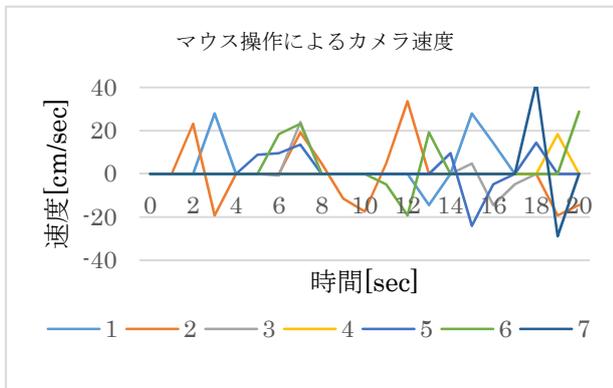
6.1. 奥側領域の自動入力改良

3.1 節で述べた奥側領域自動推定のアルゴリズムは改良の余地がある。たとえば画像中のリーディングラインを抽出すればより正確な消失点設定が可能で、目的矩形のアスペクト比も判明するので、矩形設定が容易になると考えられる。リーディングライン抽出に関する手法[17]は数多く発表されているので、その中から適したものを選択し追加実装していきたい。現在は入力画像内に消失点が存在する場合のみしか対応できないが、リーディングラインが抽出できれば、消失点が画像領域外に存在するような入力画像でも対応できると考えられる。

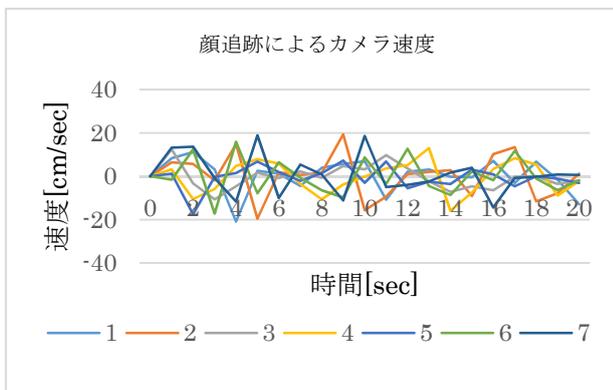
また、正確さという面ではユーザによる手動での矩形入力のが最も優れているが、自動推定は簡便さというメリットに加えて、図 8 最下行のような一見ユーザがどこに矩形入力してよいかわからないような画像に対しても、有効にはたらく可能性がある。

6.2. 3 点透視図への対応

人がウォークスルーしたいと感じるような画像の種類は限られていると予想される。例えば、山頂から一望したような見通しのよい風景はその可能性が高く、一方、人の顔写真に関してはその可能性は低いと考えられる。この仮定に従えば、1 点透視図と 2 点透視図に加えて 3 点透視図にも対応できれば、本システムの入力画像として想定される画像にはほぼ対応できる。



(a): マウス操作によるウォークスルー時のカメラ速度



(b): 顔追跡によるウォークスルー時のカメラ速度

図 11: 異種のインタラクション方法によるカメラ速度測定結果

3点透視図の場合、2点透視の場合と同様に直方体を2つ生成し、推定された3つ目の消失点に付随するリーディングラインに沿って直方体のスケールを調整し歪ませることで、第3の消失点方向にもユーザがウォークスルー可能な空間を創出することを考えている。

6.3. 円柱へのテクスチャリング



図 12: 直方体へのテクスチャリングになじまない画像の例

図 12 のように明確なリーディングラインをもたず、消失点から均一に物体が分布している画像の場合、テクスチャのマッピング対象を直方体ではなく円柱にした方が自然な見目が得られる。2章で述べたが、入力画像がもつ構図の特性に応じた3次元構成は TIP の拡張手法[18]として提案されている。入力画像に応じて使い分けられるように対応していきたい。ここで、円柱にテクスチャリング対象を切り換えた場合でも3章で説明した手法とほぼ同じフローで3次元構成が行えるので、あくまでも簡便さの利点を失っていないことに注意されたい。

6.4. 障害物の除去および配置

入力画像中に前景オブジェクトが存在する場合、そのまま本手法を適用してしまうと、壁面のテクスチャとして貼り付けられ、不自然な結果が得られてしまう。それを避けるために、3.1節の奥側領域推定に先行して、ユーザが前景オブジェクトを選択し、画像上から除去したのち、別レイヤにて自動的に再配置する。前景オブジェクトの位置と大きさは、ユーザ入力および、この後設定される奥側の矩形領域から相対的に決定する。ただし、このままではユーザ入力が必要なため、本システムの理念の柱である簡便かつ即座というメリットが喪失してしまう。そこで、視線追跡によりユーザが注視している付近の画素値を解析することで前景オブジェクトを抜き出す、半自動的な障害物除去手法を検討している。

6.5. 注視範囲以外のぼかし

結果画像(図8)からわかるように、表示対象のすべての部分にフォーカスが当たっており、どの部分のテクスチャもくっきりと表示されている。これは人の目の生理的性質とはかけ離れたものであり、ユーザに違和感を与えてしまう。そこで、ユーザの注視点付近以外にブラーをかけることで、視線に応じた疑似的な被写界深度の変更を検討している。また奥行き知覚の指標の1つである空気遠近法を再現するため、視点からの距離に応じて色相を変化させることも考えている。

6.6. ジャイロセンサの併用

顔検出を行うアプリケーションが最もユーザを興ざめさせてしまうことの1つは、ユーザの顔が認識されず、正しく機能しないことである。本システムは屋外など様々な場所で用いられることを想定しているため、照明環境によっては顔認識がされないことも考えられる。最近では、本手法の想定環境としているスマホ等の携帯端末には、回転角速度を測定するセンサであるジャイロセンサや加速度センサを搭載している機種が多い。この各種センサを用いて、顔追跡ができない状況でも端末の傾きや加速度からユーザ頭部位置を割り出し、動作を継続可能にすることを考えている。

6.7. 直方体内部における照明環境の改良

バーチャル空間内に、カメラの位置および挙動と同期させた点光源を配置し、ユーザの視点に応じて直方体内部の陰影が変化するように試みた。しかし4章で述べた実行環境のもと、表1の画像8を入力として実行したところ、平面感が強調され、かえって立体感を損なう結果となった。

現時点での光源配置では、入力画像のテクスチャに関して考慮していないため、リアリティを逆に削ぐ結果になったと考えられる。そこで、入力画像として屋内画像を用いる場合、画像の輝度分布から光源分布を推定し、推定結果に応じて種類、位置および強度を決定した光源を直方体内に配置することで、画像の元となった風景の本来もつ照明環境を再現することを検討している。ここに動作環境のハード性能をそれほど要求しない、擬似的にテクスチャ表面に凸凹をつけるバンプマッピングなどをテクスチャ生成時に適用することで、陰影による立体感を増幅させたいと考えている。

6.8. 聴覚と触覚による奥行き効果

ここまで、ユーザに臨場感を与える方策を主に視覚や前庭感覚に基づいて言及してきた。1章で述べたように、臨場感を高める要件である空間要素、身体要素、時間要素を満足する対象は視覚や前庭感覚だけではなく、外界からの刺激を受容する人間の五感全てである。ここで、スマホのような携帯端末で視覚や前庭感覚以外にどの感覚を発生させられるかを考えたとき、現時点で実現できるのは聴覚と簡単な触覚であると考えられる。

旭ら(2013)[19]が示すような、軽薄短小なデバイス上における、音による奥行き効果生成手法を本システムに組み入れることで、臨場感を増強させたユーザ体験を提供できると考えている。さらに、携帯端末のバイブレーション機能を利用することで、簡易的な触覚を再現することも可能であると考えられる。

謝辞

消失点の推定手法についてご指導いただいた山梨大学大学院総合研究部の茅暁陽教授, 実装面でお世話になった慶應義塾大学藤代研究室の中山雅紀氏に謝意を表す。また, 検証実験に協力していただいた実験参加者の皆様にお礼申し上げる。さらに, 論文の改善に対し有益な助言をいただいた査読者の皆様に感謝申し上げます。

本研究の一部は, 科研費挑戦的萌芽研究 25540045, 15K12034 の支援により実施された。

参考文献

- [1] 井阪建, 藤代一成: “IFWB: 没入型映像インスタントメーカ”, 芸術科学会 NICOGRAPH2014, 2014.
- [2] Elizabeth Bell: *Theories of Performance*, SAGE Publications, 2008.
- [3] 山本奏, 田中秀典, 安藤慎吾, 片山淳, 筒口拳: “Visual SyncAR: 映像に同期して情報を重畳表示する映像同期型 AR 技術”, 画像電子学会誌, vol.43, no.3, pp. 397-403, 2014.
- [4] 安藤広志, カラン明子, Norberto Eiji Nawa, 西野由利恵, Juan Liu, 和田充史, 坂野雄一: “臨場感の知覚認知メカニズムと評価技術”, 情報通信研究機構季報, vol.56, no.1, pp. 157-165, 2010.
- [5] Brian Rogers and Maureen Graham: “Motion parallax as an independent cue for depth perception,” *Perception*, vol. 8, pp. 125-134, 1979.
- [6] Tsuyoshi Suenaga, Yoshio Matsumoto and Tsukasa Ogasawara: “3D display based on motion parallax using non-contact 3D measurement of head position,” *OZCHI '05 Proceedings of the 17th Australia Conference on Computer-Human Interaction: Citizens Online: Considerations for Today and the Future*, pp. 1-4, 2005.
- [7] 末永剛: 視点位置計測に基づく運動視差 3 次元ディスプレイの開発と立体感の評価, 奈良先端科学技術大学院大学情報科学研究科博士論文, 2008.
- [8] Youichi Horry, Ken-ichi Anjyo and Kiyoshi Arai: “Tour into the Picture: Using a spidery mesh interface to make animation from a single image,” *ACM SIGGRAPH 97 Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 225-232, 1997.
- [9] Satoshi Iizuka, Yoshihiro Kanamori, Jun Mitani and Yukio Fukui: “Efficiently modeling 3D scenes from a single image,” *IEEE Computer Graphics and Applications*, vol. 32, no. 6, pp. 18-25, 2012.
- [10] Kevin Boulanger, Kadi Bouatouch and Sumanta Patanaik: “ATIP: A tool for 3D navigation inside a single image with automatic camera calibration,” *Theory and Practice of Computer Graphics 2006*, pp. 71-79, 2006.
- [11] Canny John: “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, 1986.
- [12] Paul Viola and Michael Jones: “Robust real-time face detection,” *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.
- [13] OpenCV2.3: <http://sourceforge.net/projects/opencv-library/files/opencv-win/2.3.1>
- [14] WPThemeDesigner.com: GATAG | フリー絵画・版画素材集 <http://free-artwork.s.gatag.net>
- [15] 瀬川かおり, 氏家弘裕, 岡嶋克典, 斎田真也: “オブティカルフローによる自己移動速度知覚に距離手がかりが及ぼす影響”, 日本バーチャルリアリティ学会論文誌, vol. 8, no. 1, pp. 111-117, 2003.
- [16] 加藤典之, 金子寛彦: “歩行応答と知覚応答による視覚誘導性自己直線運動の知覚速度の測定: 刺激形状の効果に関する検討”, 光学, vol. 34, no. 11, pp. 597-605, 2005.
- [17] 後藤悠汰, 澤田友哉, 豊浦正広, 茅暁陽, 行場次朗: “リーディングラインを考慮した顕著性マップの作成”, *Visual Computing/グラフィクスと CAD 合同シンポジウム 2014*, 2014.
- [18] Hyung Woo Kang¹, Soon Hyoung Pyo, Ken-ichi Anjyo and Sung Yong Shin: “Tour Into the Picture using a Vanishing Line and its Extension to Panoramic Images,” *Computer Graphics Forum*, vol. 20, no. 3, pp. 132-141, 2001.
- [19] 旭克也, 伊藤智也, 伊藤弘樹, 菊池司: “タブレット端末における視覚情報と聴覚情報の奥行き知覚増幅法”, 芸術科学会論文誌, vol.13, no.1, pp. 34-44, 2013.

井阪 建 (いさか たける)



2014 年慶應義塾大学理工学部情報工学科卒業。工学士。現在, 同大学大学院理工学研究科在学中。立体感や臨場感を与えるようなグラフィクスの研究に従事。情報処理学会学生会員。

藤代 一成 (ふじしろ いっせい) 芸術科学会会員



1985 年筑波大学大学院博士課程工学研究科電子・情報工学専攻修士号取得退学。1988 年理学博士(東京大学)。東京大学助手, 筑波大学助手・講師, お茶の水女子大学助教授・教授, 東北大学教授を経て, 2009 年より慶應義塾大学理工学部情報工学科教授, 現在に至る。ビジュアルコンピューティング, 特にボリュームグラフィクスや可視化ライフサイクル支援, 複合現実型並置化と多感覚情報呈示による知的環境メディアの実現に関する研究に従事。本会では理事, 副会長を歴任し, 現在評議員。画像電子学会 VC 委員会委員長, IEEE Computer Society, ACM, Eurographics, CGS, 画像電子学会, 情報処理学会, 日本計算工学会, 日本バーチャルリアリティ学会, 映像情報メディア学会, 電子情報通信学会各会員。