

DVF：スケールの個人差を考慮した仮想シーン知覚像の半自動再現

星川 潤¹⁾ 藤代 一成²⁾(正会員)

1) 慶應義塾大学 大学院理工学研究科 2) 慶應義塾大学 理工学部

DVF: Semiautomatic Reproduction of Perceptual Images of Virtual Scenes Considering Individual Differences in Scale

Jun Hoshikawa¹⁾ Issei Fujishiro²⁾

1) Graduate School of Science and Technology, Keio University

2) Faculty of Science and Technology, Keio University

{jun.hoshikawa, fuji}@fj.ics.keio.ac.jp

概要

旅行先などで、印象的な情景を写真に収めた際、写真に記録された風景と自身が肉眼で捉えた風景との間に大きな違和感を覚えることがある。この現象は、カメラの投影モデルと人間の視知覚の特性の違いが原因のひとつであると考えられる。一般的なカメラが風景全体を平等に捉えるのに対し、人間は風景のなかで興味ある対象を部分的に強調しながら捉える。加えて、仮想シーンでは正確な大きさ知覚が現実シーンよりも難しいことが知られており、仮想シーン上で得た主観的な印象や感動を他者に効果的に伝えることは容易ではない。そこで本論文では、ユーザの注目部分を適度に拡大処理した画像を自動出力するシステム Digitus ViewFinder (DVF) を提案する。注目部分の特定にジェスチャインタフェースを用いることで、ユーザは没入感を損なわずに直観的で正確に範囲を選択できる。そして、試行とともにユーザの拡大率における個人差を学習することで、段階的に提案の精度が向上するサジェスティブインタフェースを実現する。実際にユーザ評価実験から、DVF システムの有意性が確認された。

Abstract

When taking a photo of impressive scene, we sometimes feel significant discrepancy between the scene we perceived and that captured by the camera. One of the factors that cause this problem lies in the difference between the characteristic of human visual perception and that of the camera's projection model. Human beings perceive scenes by enlarging an area of interest (AOI), whereas an ordinal camera captures the entire scene evenly. In addition, it is known that accurate size perception is more difficult in virtual scenes than in real scenes, thus it is not easy to share subjective impressions obtained in the virtual space effectively with others. In this paper, therefore, we propose Digitus View Finder (DVF), a system that automatically outputs images close to the scene the user perceives. The user is allowed to operate this system immersively, intuitively, and accurately using a gesture interface to identify their own AOI. The suggestive interface progressively improves the accuracy of scale factor by learning individual differences practically. The effectiveness of the DVF system was confirmed empirically through a user study.

1 序論

旅行などの際、写真に記録した風景が実際に肉眼で見た風景から大きく異なると感じることもある。このような現象が起こる原因の一つとして、大きさの恒常性が挙げられる [1]。大きさの恒常性は、距離が変化しても知覚するオブジェクトの大きさは変化しにくいという人間の認知メカニズムをさす。カメラは透視投影モデルの原理に従うため、オブジェクトの大きさが距離に反比例して小さくなるのに対し、人間は大きさの恒常性によって、離れたオブジェクトに注目する際、本来よりも大きく知覚する傾向がある。この特性の違いにより、同じシーンであっても、我々はカメラで撮影した風景と自分が主観的に捉えた風景との間に違和感を覚える。

認知心理学の分野では、前者の透視投影された像を「網膜像」、後者の拡大された主観的な像を「知覚像」とよぶ。知覚像の特性を応用し、臨場感あふれる絵画やアニメーションを制作したケース [2] は多数存在するが、その設計は個人のスキルに依存しており、網膜像を自動的に知覚像に変換するシステムはまだ知られていない。

知覚像の再現は、現実シーンのみならず仮想シーンにおいても重要な課題である。仮想シーン上では、距離知覚および大きさ知覚が環境に大きく依存しているため [3][4][5]、知覚像と網膜像との間の乖離の問題は現実シーンのそれよりも複雑である。よって、仮想シーンで得られる多様な体験を他者と効果的に共有することは難しい。このような課題に対し我々は、ユーザの注目部分を強調した画像を半自動的に生成するシステム Digitus ViewFinder (DVF) を提案し、仮想シーンにおけるユーザの知覚像の効果的な再現を目指す。著者らの先行研究 [6, 7] によれば、知覚像を効果的に画像として表現するためには、強調の度合いに関する個人差を考慮する必要がある。しかし、これには多くの要因が潜在していると考えられ、ユーザごとに最適な拡大率の画像を即時に出力することは容易ではない。そこで本論文では、複数の拡大画像を提示するサジェスティブインタフェースを導入し、試行とともにユーザの個人差を学習するシステムを提案する。これにより、ユーザに対する拡大画像の提案の精度が段階的に向上することを目指す。実際に複数シーン上で DVF を用いて撮影する評価実験を行うことで、システムの有意性を確認した。

本論文は次章以降以下のように構成されている。次章では、DVF がもつ、人間の視知覚特性とジェスチャインタフェースという二つの側面に関する関連研究を示し、3章では DVF のインタフェースとシステムについて説明する。4章では、DVF の性能を確認するために実施した評価実験の流れを説明し、5章に実験データの分析結果とそこから得られた考察を示す。6章では、本論文の結論を述べ、今後取り組むべき課題に言及する。

なお、本研究の概要や DVF が実際に動作する様子については、添付ビデオを確認してほしい。

2 関連研究

提案手法は大きく二つの技術的側面に分けられる。第一に、人間が知覚した情景を効果的に再現するという知覚特性の側面。そして第二に、ジェスチャを利用するという、インタフェースとしての操作性の側面である。本章では、各領域を対象とした先行研究について説明し、本研究の位置づけを明確にする。

2.1 視知覚特性の再現

人間の視知覚と透視投影モデルの幾何学特性の違いは、これまで数多くの研究で分析されてきた。大きさの恒常性により、人間は注目するオブジェクトを本来のサイズよりも大きく認識した知覚像を結ぶ。知覚像を画像で再現するための有効な手段の一つとして、シーン上のオブジェクトの適度な拡大処理が考えられる。実際、長田ら [8] は、知覚される大きさと観察物の距離の関係を示す拡大率関数を推定した。このようなオブジェクトの大きさの錯覚は、仮想シーン上では様々な要因によりさらに複雑化する。Plumert ら [9] は現実環境と CAVE における距離知覚の違いを調査し、Murgia ら [5] は同環境における奥行の手がかりと距離知覚の関係について調査した。これらの研究では、人間は現実環境よりも仮想環境の方が距離を過小評価しやすいという結論が得られている。Head Mounted Display (HMD) 環境においても、CAVE と同様に正確な距離知覚が難しい [10]。その原因として、レンズによる歪曲効果や、眼の輻輳距離と焦点距離の不一致などが挙げられる [11][12]。Liu ら [13] は、HMD における距離知覚を現実に近いものにするために、オブジェクトとの距離に応じて視野角を変化させる投影手法を提案した。このように、仮想シーン上では距離知覚が不正確であるため、スクリーンが提示す

る映像とユーザが捉えた知覚像は大きく乖離している。

そこで本論文では、仮想シーン上のオブジェクトを直接拡大させたシーンを撮影することで、知覚像に近いシーン画像を半自動生成するシステムを提案する。

2.2 ジェスチャインタフェースの応用

ジェスチャインタフェースとは、人間にとって最も操作性に富んだ器官である手を介したインタフェースである。デバイス依存性が低く、没入感も損なわないといったメリットが存在しているため、ユビキタスコンピューティングの実現に最も有効な手段の一つであると考えられている。なかでも遠距離のオブジェクト選択を目的としたジェスチャインタフェースは数多く研究されてきた [14][15]。レーザーポインタ [16] や Go-Go システム [17], WIM [18] など様々な手法が検討されてきたが、これら多くの手法はオブジェクト単体を選択することを目標としており、複数オブジェクトの同時選択や空間の範囲指定を目指した手法は多くない。古山ら [19] は、指フレームという両手の親指と人差し指で矩形枠を形成するジェスチャを用いて、任意の構図で写真撮影するインタフェースを提案したが、物理的制約として指に外部センサを取り付ける必要がある。Eureka Works [20] は、仮想シーン上で指フレームを使って範囲的な文字認識や QR コード認識を可能にするアプリケーション VRHandsFrame を提案したが、インタフェースとして両手コントローラを必要とする。

提案手法では、深度カメラによるハンドトラッキング技術を応用することで、手をそのまま用いた指フレームによる範囲指定を実現する。

2.3 本研究の位置づけ

2.1 節で説明したとおり、知覚像と距離の相関を対象とした解析研究が進む一方で、シーンの特性やユーザの個人差を考慮した知覚像の生成システムに焦点を当てた研究は未だ数多くない。このようなシステムを実現するためには、人間が注目する部分を特定し、その部分を知覚した大きさに拡大処理する必要がある。本論文では、知覚像に対する実験と分析によって得られた結果を応用し、簡単なインタフェースでより正確に知覚像を出力するアプリケーションを提案する。直観的かつ手軽にユーザの注目部分を特定するために、2.2 節で説明したジェスチャインタフェースを利用する。図 1 のように、注目範囲を指フレームのジェスチャで指定しながら撮影する

インタフェースを介して、仮想シーンにおけるユーザの知覚像を半自動的に生成することを目指す。

3 システム

DVF の処理工程の概略を図 2 に示す。その特徴はユーザの注目部分の特定と拡大という二種類の処理にあり、前者はジェスチャ認識を、後者はジオメトリ変化を用いている。そして、注目部分の拡大率に関する個人差を学習することで、ユーザごとに適した知覚像を提案できるサジェスティブインタフェースを実現する。

開発に用いたデスクトップ環境は、CPU: Intel Core i7-10875H 3.50 GHz, RAM: 16.00 GB, GPU: NVIDIA GeForce RTX 2080 SUPER Max-Q であり、システムは Unity 2019.4.21f 上で実装した。VR デバイスは HMD である VIVE Pro Eye を用いており、ジェスチャの検出には HTC が提供しているハンドトラッキング機能 VIVE Hand Tracking SDK 1.0.0 を利用した。

3.1 インタフェースの概要

DVF は、HMD を装着したユーザが指フレームで注目部分を囲いながら撮影するインタフェースをもつ。撮影

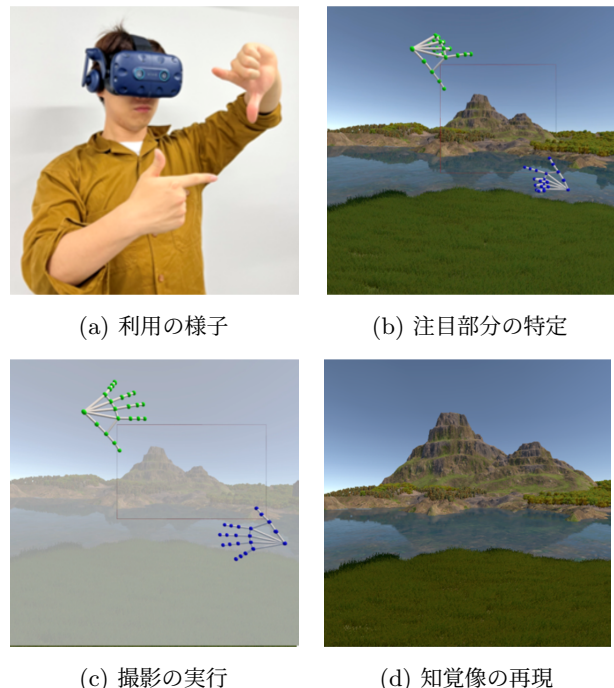


図 1: DVF の使用例。指フレームのジェスチャ (a) を用いて注目部分を特定し (b), 「パー」のジェスチャで撮影を実行 (c)。注目部分を拡大することで、ユーザの知覚像に近いスケール画像が自動的に出力される (d)。詳細は添付ビデオを参照。

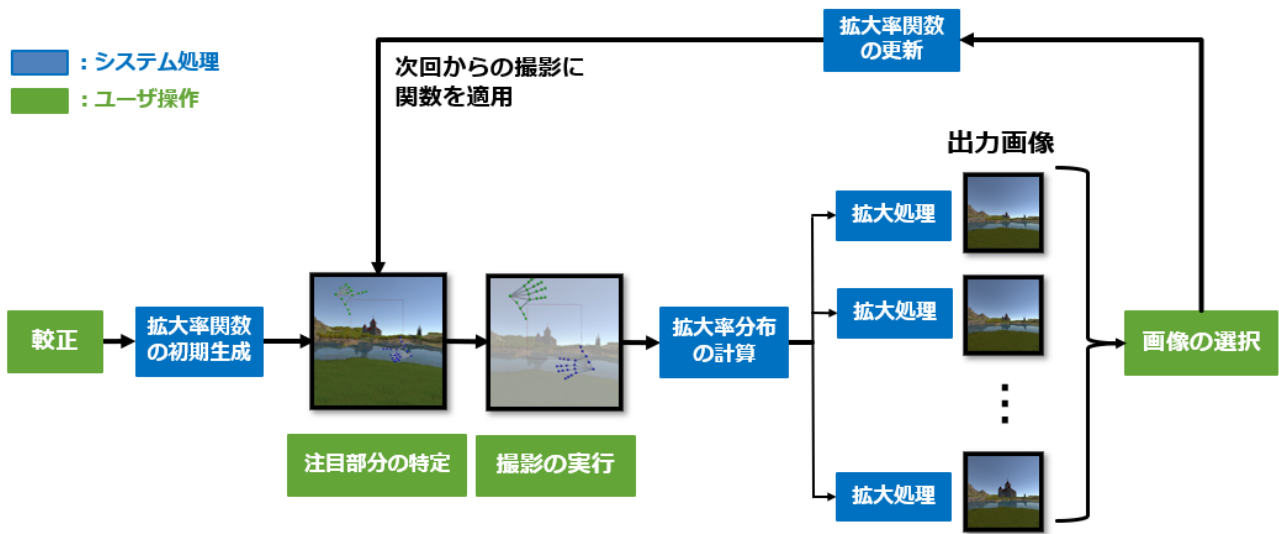


図 2: DVF における処理の流れ

の際に注目部分に位置するオブジェクトのスケールを変化させながらスクリーンショットを行うことで、ユーザの視野全体を含みつつ注目部分が適度に強調された画像を出力する。

先行研究 [6, 7] によれば、知覚像の再現に必要な強調の度合いには個人差が存在しており、ユーザごとに最適な拡大率の画像を即時に出力することは容易ではない。そこで、我々は複数枚の拡大画像をユーザに提示し、そのなかから自身の知覚像に最も近い画像を選択させるサジェスティブインタフェースを導入する。操作は添付ビデオを参照してほしい。そのうえで、撮影と選択の試行を繰り返す間にユーザの最適拡大率を学習することで、段階的に提案の精度が向上するシステムを提案する。

3.2 較正

撮影の準備として、効果的に視界全体を画像として出力するために、ユーザの有効視野角と視野のアスペクト比を較正する。ここで、有効視野角とは人間が認知可能な視野角であり、参加者が捉えた風景全体を一枚の画像に出力する際に最適な画角のことをさす。Vive Pro Eye が表示できる視野角は 110 度であるが、既存研究 [21] によれば、表示画面のうちユーザが認識できる範囲は限定的である。そこで、DVF では撮影用のカメラはスクリーン投影用のカメラとは別に設定し、撮影用カメラの画角にユーザの有効視野角と視野のアスペクト比を適用させ

ることで、ユーザが実際に見た情景の全体感の忠実な再現を試みた。また、撮影用カメラは仮想シーンにおけるユーザの両眼の中間に置くことで、両眼で捉えた像を代表する一枚の知覚像を出力することを目指した。

3.3 注目部分の特定

DVF では、まずユーザがシーンにおける注目範囲を特定するステップから始まる。同一のシーンを見ていても、ユーザによって注目部分は異なるため、ユーザごとに知覚像を効果的に再現するためには、個々の注目範囲を正確に抽出しなければならない。一般的にそのような注目範囲の特定には、視線追跡が有効であると考えられるが、人間の視線は不安定であるうえに出力はポイントデータのみであるため、視線データのみによる範囲特定は非効率である。そこで我々は、シーン中の注目部分を指フレームのポーズで囲みながら撮影するインタフェースを開発した。指フレームとは一般的にカメラマンや画家が作品の構図を決定する際に用いるジェスチャであり、両手の親指と人差し指で矩形枠をつくるポーズをさす。このような馴染み深いジェスチャを仮想シーンに導入することで、多くのユーザは直観的かつ手軽な操作で自身の注目範囲を正確に指定することができる。

提案手法では、指フレームのジェスチャが検知されると矩形枠が表示され、ユーザはその矩形枠をリアルタイムに動かすことができる。そして左手を「パー」のジェ

スチャにした瞬間、撮影が実行される。

3.4 注目部分の拡大

撮影が実行されると、矩形枠に含まれたオブジェクトが拡大された状態の風景が画像として出力される。なお、矩形枠内に存在するオブジェクトが複数の場合でも、それらは全て拡大処理される。オブジェクトが矩形枠に含まれているかどうかの内外判定は、オブジェクトごとにバウンディングボックスを割り当て、矩形枠から生成される視錐台との交差判定によって計算する。

拡大処理はオブジェクトのジオメトリの直接操作で実現しており、これは現実シーンと異なりセグメンテーションを必要としない。オブジェクトを拡大し撮影してから、画像を保存するという一連の処理を、HMD スクリーンに対するレンダリングと並行させることで、撮影中のユーザに対するシーンの提示動作が変化しないまま、複数の異なる拡大率画像を同時に出力することができる。

先行研究 [6, 7] の実験結果を用いて、我々はユーザの視野角 e 、オブジェクトまでの距離 d 、そして注目部分のオブジェクト数 n の 3 種類のパラメタから構成される拡大率関数 f を推定した。

$$f(e, d, n) = \delta(\mu - e) + \kappa d + c(n) \quad (1)$$

$$c(n) = \alpha - \beta(n - 1) \quad (2)$$

実験結果に対して式 (1) を最小二乗法によって当てはめたところ、 $\alpha = -6.1 \times 10^{-2}$ 、 $\beta = 1.1 \times 10^{-2}$ 、 $\delta = 1.0 \times 10^{-2}$ 、 $\mu = 6.7 \times 10$ 、 $\kappa = 6.0 \times 10^{-3}$ となった。しかし、式 (1) の二乗平均平方根誤差は 0.247、決定係数は 0.368 であり、モデルとしての精度は高いといえない。そのような原因として、拡大率に関する個人差が大きいこと、そして同じユーザでも拡大率にばらつきがあることが考えられる。そこで、我々はユーザの拡大率を正規確率分布と仮定し、出力画像群の分布を徐々に各ユーザ固有の分布に近づけることで、サジェスションの精度が段階的に向上するシステムを実現することにした。

3.5 個人差の学習

DVF が提示する画像の集合とその画像一枚を各々 Y と S_y とおき、さらに拡大率関数が推定するユーザの最適拡大率を S_s とおいたとき、 Y と S_s はそれぞれは以下

のような式で示される：

$$Y = \{S_y | S_s - \sigma \leq S_y \leq S_s + \sigma\} \quad (3)$$

$$S_s = f(e, d, n) + S_m \quad (4)$$

ユーザが提案画像群 Y から知覚像に最も近い拡大率画像を選択する作業を繰り返すなかで、固有拡大率平均値 S_m と固有拡大率偏差 σ はそれらの拡大率を 95% 信頼区間で包含するよう更新される。これにより、ユーザが撮影と画像選択の試行を重ねるにつれ、DVF のサジェスションの精度は向上する。

4 評価実験

DVF のシステムの性能と拡大率関数の精度を評価するために、14 名の実験参加者 A から N を対象とする新たな実験を実施した。実験の参加者は著者が所属する大学の大学院生および学部生 (20 代の男子 11 名、女子 3 名) であり、そのうち 2 名は VR 未経験者であった。

4.1 手続き

実験の手続き全体を図 3 に示す。参加者には、DVF を実際に利用した撮影タスクと、拡大率が異なる画像群から自身の知覚像に最も近いものを選択する評価タスクの 2 種類が与えられた。これらのタスクを、図 4 に示す 4 種類のシーンで行い、シーンの切替え時には評価タスクの結果をもとに拡大率関数を更新した。

提示するシーンとその順番は図 4 に示すとおりである。シーン 1 と 2 は深度キューが少なく、主にユーザ固有の拡大率分布をおおまかに推定することを目的としたプリミティブなシーンである。一方、シーン 3 と 4 はコンテキストが豊富なアウトドアシーンとインドアシーンであり、実際に仮想シーンの撮影シナリオを想定している。なお、シーン 3 には複数のランドマークが映っていることに注意してほしい。

4.2 事前準備

評価実験の事前準備として、参加者に対し有効視野角と視野のアスペクト比の較正を行った。計測の手続きは以下のとおりである。まず、HMD を装着した参加者に対して、図 5 のような十字型にテープが重なったオブジェクトを視野の中央に提示した。計測が開始すると、横方向のテープを時間経過とともに少しずつ長くした。そして、参加者が中心の赤い点に注目した状態で、テープが視野の端に達するタイミングを計測した。この計測

を横と縦の方向でそれぞれ3度実施することで、ユーザ固有の視野角と視野のアスペクト比の値を取得した。そして、それらの値の平均値を、以後のシーンにおける撮影用カメラの画角に適用した。

4.3 撮影タスク

タスクに入る前に、参加者に DVF の具体的な仕様が説明し、実際に自身が気に入った範囲を矩形枠で囲みながら撮影させた。参加者は提示されるシーンのなかに点在する3つ以上のランドマークを対象に、各シーンにつ

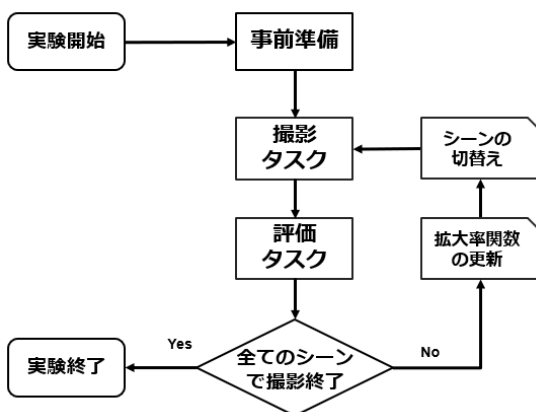


図 3: 評価実験の流れ。詳細は添付ビデオを参照。

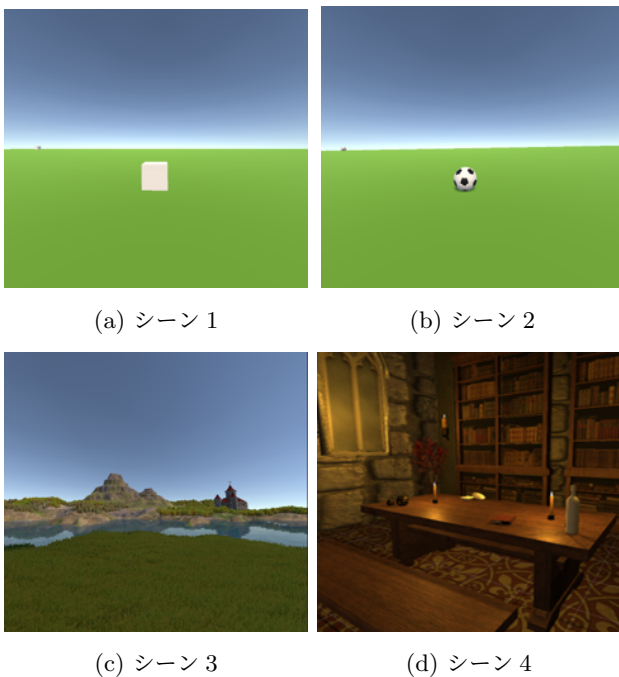


図 4: 評価実験に用いた4種類のシーン

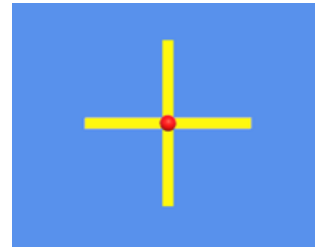


図 5: ユーザの視野の較正に用いたオブジェクト

き3枚撮影した。特にシーン1,2では正確な較正を行うために、ランドマークとユーザ間の距離はばらつきが大きくなるよう設定されている。撮影の際、参加者には自身が捉えた景色をよく記憶するよう指示し、撮り終わり次第評価タスクに移行させた。

4.4 評価タスク

撮影する際、DVFで囲んだオブジェクトが異なる拡大率で拡大された画像が自動で複数枚保存される。拡大率関数によって出力された11枚の提案画像群Yに加え、先行研究[6, 7]と同様、評価に必要な1.0~2.5倍の範囲で拡大率が0.1ずつ変化する画像も出力された。参加者にはこれらの画像のなかから、自身が感じた情景に最も近い拡大率 S_b の画像を回答させ、その満足度を7段階で評価させた。その後、拡大率関数が推定するユーザの最適拡大率 S_s の画像と、拡大処理されていない元画像を提示し、それぞれの満足度を7段階で評価させた。

この評価タスクでは、拡大率に対するバイアスを生じさせないため、参加者に最初に提示する画像はランダムに選択した。したがって、参加者は自身が見ている画像の拡大率の数値を確認することはなかった。

4.5 拡大率関数の更新

DVFで撮影した3枚の画像に対する評価タスクが終了後、シーンの切替えとともに拡大率関数を更新した。DVFはユーザが選択した拡大率のログを取得し、拡大率関数の固有拡大率平均値 S_m と固有拡大率偏差 σ を更新する。なお、 S_m と σ の初期値は $S_m = 1.56$, $\sigma = 0.27$ であり、先行研究[6, 7]の実験結果の拡大率の平均値と標準偏差を用いた。

5 結果と考察

本章では、DVFが出力する拡大画像の満足度および拡大率関数の精度に関して分析した結果を示す。また、5.3節で述べるとおり、実験参加者のうち、深度キュー

の解釈能力が特に高い 2 名の実験結果は外れ値として扱った。

5.1 拡大率関数の精度

本節では、2 種類の指標を用いて拡大率関数の精度を評価する。図 7 は各シーンにおける、個人差に対する学習前拡大率 S_i と学習後拡大率 S_s の、最適拡大率 S_b からの平均二乗誤差を示している。 S_i は、個人差を学習せず、 S_m を初期値のまま更新しなかった場合に拡大率関数から出力される拡大率である。

図 7 から、個人差がまだ学習されていないシーン 1 以外の全てのシーンで、 S_s は S_i よりも誤差が小さいことがわかる。これは、個人差の学習によって DVF がよりユーザの理想に近い拡大率画像を出力できていることを示している。

なお、 S_s と S_i のどちらにおいても、シーン 1,2 よりシーン 3,4 の方が誤差が少ない。この原因として、ユーザとオブジェクト間の距離が大きく関係していると考えられる。シーン 1,2 はユーザ固有の拡大率分布の較正のため、撮影対象物の距離が分散しているのに対して、シーン 3 は全てのオブジェクトが遠方に、シーン 4 は全て近方に位置しており、前者に比べ距離が分散していない。既存研究 [10] でも述べられているとおり、距離は知覚像の拡大率に大きく影響する要因であるため、全てのオブジェクトが身近な距離に位置するシーン 4 は特に拡大率の値のぶれが少なく、誤差が最も小さくなっている。

図 8 は、各シーンにおける提案適正率 E を示している。 E は、ユーザが選択した最適拡大率 S_b の画像が提

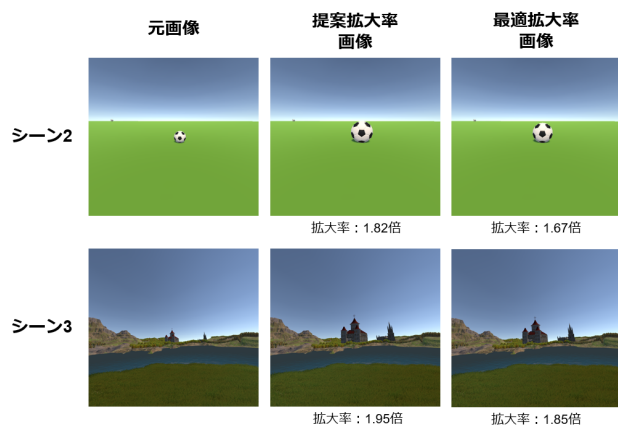


図 6: 評価実験で出力された拡大画像の例

案画像群 Y の分布に含まれる確率を表す。DVF の提案画像群は式 (3) のとおり、 S_s を平均値、 σ を標準偏差とした正規分布をもとに生成されており、図 8 はその分布の 65% 信頼区間と 95% 信頼区間それぞれにおける E を示している。図 8 から、65% と 95% の両方の信頼区間において、 E がシーンの切替えとともに高くなっていることがわかる。特に 95% 信頼区間におけるシーン 4 の E は 100% という結果が得られており、信頼区間 65% においても 84.6% という高い数値が得られた。これらの結果から、拡大率関数とその更新アルゴリズムによって、提案画像群の分布がユーザ固有の拡大率分布に段階的に近づくことが示された。

5.2 満足度の評価

本節では、DVF が出力する画像に対するユーザの満足度を分析した結果を示す。図 6 に参加者 D の評価実験で実際に出力された画像の例を示し、図 9 に各シーンにおけるユーザ全体の満足度評価の結果を示す。図 9 では、緑色は元画像、橙色は拡大率関数によって算出され

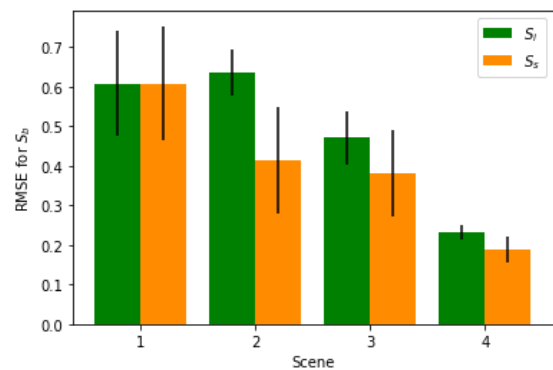


図 7: シーンごとにおける最適拡大率との平均二乗誤差

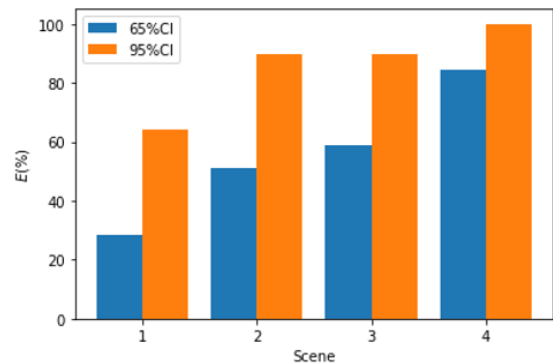


図 8: ユーザが選択した画像が提案拡大率の画像群に含まれる確率

た提案拡大率 S_s の画像，青色はユーザが自身の知覚像に最も近いと選択した最適拡大率 S_b の画像の満足度を表している。

元画像と他の画像を比べると，元画像の満足度は低くシーンの切替えによる数値の向上が見られない．その一方で，シーンが切り替えられるごとに S_b と S_s の満足度は向上している． S_b の満足度が向上する要因は大きく二つあると考察する．一つはユーザが HMD を介した仮想シーンの知覚に徐々に慣れたため，そしてもう一つはシーン 1,2 より 3,4 の方が情景的で印象に残りやすいためであると考えられる．実際参加者 A は「シーン 1 と 2 は無機質すぎて思い入れが特になかった」というコメントを述べていたため，シーンのコンテキストが拡大率に与える影響は小さくないといえる．

また図 9 から，シーンが切り替えられるごとに， S_s の満足度が段階的に S_b の満足度に近づいていることがわかる． S_s の満足度が向上する要因は， S_b よりも向上の勾配が大きいことから，拡大率関数の学習による効果が大きいと考えられる．

5.3 深度キューの解釈能力の個人差

本節では，外れ値に焦点を当て，本格的な空間把握の経験が拡大率に与える影響について考察する．図 10 は，外れ値として数えた 2 名の参加者 {K,I} と，その他の参加者の最適拡大率 S_b の平均値をそれぞれ示している．同図からわかるとおり，シーン 1,2 とシーン 3,4 の間の S_b の平均値の差異は，{K,I} の両名が特に大きいという結果が得られた．このような結果の原因として，我々はユーザのオブジェクトに対する深度キューの解釈能力の

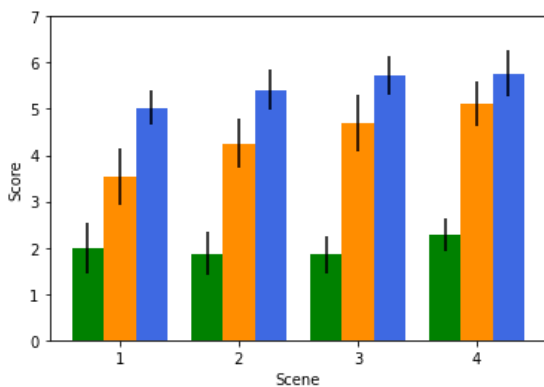


図 9: シーンごとにおける元画像 (緑) と提案拡大率画像 (橙) と最適拡大率画像 (青) の満足度評価の平均値

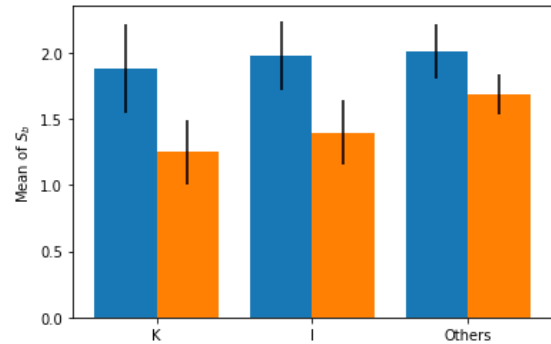


図 10: {K,I} と他の参加者の最適拡大率の平均値 (青：シーン 1,2, 橙：シーン 3,4)

違いが大きく関係していると考えられる。

{K,I} の 2 名に共通する要因として，豊富な美術製作経験を有していることが挙げられる．実際，参加者 I は実験後，「美術学校でデッサンの練習をしていた際，相対的に物体のスケールや距離を把握する練習をしており，その習慣は今も身につけている」というコメントを述べていた．シーン 3,4 はシーン 1,2 に比べ，深度キューやコンテキストが豊富であるため，{K,I} の両名は無意識的に他者よりも正確なスケールを認知し，大きさの恒常性が働きにくかったと考えられる．このようなことから，人の注目部分の強調の度合いはその人の深度キューの解釈能力に大きく依存しており，美術製作などでその能力を自然に身に着けた人とそうでない人の中には，拡大率に関する有意差が存在すると推測できる．

6 結論と今後の課題

本論文では，仮想シーン上における知覚像を画像として効果的に再現するために，ジェスチャインタフェースを介してユーザの注目部分を適度に拡大処理するシステム DVF を提案した．また，ユーザの個人差を学習することで，段階的に提案の精度が向上するサジェスティブインタフェースを実現した．評価実験の結果から，試行を重ねるとともに，DVF の出力画像の満足度と拡大率関数の精度が向上することが確認できた．

本手法について，今後取り組むべき課題を以下に提示する．

6.1 拡大率の決定要因の探索

拡大率を決定する要因として，先行研究 [6, 7] ではオブジェクトまでの距離や注目部分に存在するオブジェクトの個数などが挙げられたが，今回の評価実験により，

シーンに対する印象や感動の度合い、そして美術製作の経験の有無などといった要素も拡大率に影響し得ることがわかった。今後も評価実験を通じて、拡大率に影響するシーンやユーザの特性を分析し、その詳細化を図っていききたい。実際に今回の実験結果では、図7の通り S_s の誤差が著しく小さくなっていることから、DVF の学習は似たシーンにおいて有効であると推測できる。そのような仮説を確認するために、類似度が高いシーン間とそうでないシーン間での学習勾配を比較する追加実験を実施したい。また、出力画像の提示環境と拡大率の間の相関についても比較実験を追加実施し、提示するディスプレイのサイズやユーザまでの距離といったパラメタを対象とした分析調査を進めたいと考えている。

6.2 シーンに依存しない拡大手法の実装

現在の DVF の拡大システムは、シーンやオブジェクトの仕様に大きく依存している。オブジェクトが密になった場合、複数オブジェクトのジオメトリ拡大はシーン全体の整合性を崩す可能性が高い。これらの課題を解決するため、我々はオブジェクト空間ではなく画像空間における処理によるアプローチを考えている。ユーザが注目していない部分を適切に省略すれば、相対的に注目部分を強調することが可能になる。よって、今後は Seam Carving [22] などの画像処理技術を用いた注目部分の強調処理 [23] を検討していきたい。

6.3 現実シーンへの拡張

DVF は現在のところ仮想シーンにおける利用のみを想定しているが、今後は AR デバイスなどを利用することで現実シーンの利用にも拡張していきたい。しかし、現実シーンの撮影では手を前方にかざすと撮影された画像に手が含まれてしまうため、現在の DVF のシステムをそのまま適用することは難しい。よって今後は、視線情報のログなどから注目部分を推定する手法を検討し、AR デバイスの視線追跡システムと組み合わせることで現実シーンに適用していきたい。

6.4 スケール以外の知覚要素の再現

人間の知覚像を実現するためには、スケール以外にも色味や情景の顕著度などの要素を考慮する必要がある。特に色知覚に関して評価した研究は数多く存在するため、今後はそれらの知見も取り込んでシステムを改良していきたい。

謝辞

本研究を進めるにあたり、実装の技術面での助言を頂いた方々、評価実験に参加して下さった方々に深く感謝する。本論文の一部は、科研費挑戦的研究（開拓）20K20481 の支援により実施された。

参考文献

- [1] 上野雄宏. 大きさの恒常性の研究. 人文研究, Vol. 20, No. 2, pp. 65–150, 1968.
- [2] 植草信和. 千と千尋の神隠しを読む 40 の目. 株式会社キネマ旬報社, 2001.
- [3] Xun Luo, Robert V Kenyon, Derek G Kamper, Daniel J Sandin, and Thomas A DeFanti. On the determinants of size-constancy in a virtual environment. *International Journal of Virtual Reality*, Vol. 8, No. 1, pp. 43–51, 2009.
- [4] Jannick P Rolland, William Gibson, and Dan Ariely. Towards quantifying depth and size perception in virtual environments. *Teleoperators and Virtual Environments*, Vol. 4, No. 1, pp. 24–49, 1995.
- [5] Alessio Murgia and Paul M Sharkey. Estimation of distances in virtual environments using size constancy. *International Journal of Virtual Reality*, Vol. 8, No. 1, pp. 67–74, 2009.
- [6] Jun Hoshikawa and Issei Fujishiro. DVF: Toward semiautomatic composition of perceptual images of a virtual scene through hand gesture interface. In *2022 International Conference on Cyberworlds*, pp. 169–170, 2022.
- [7] 星川 潤, 藤代 一成. DVF: ハンドジェスチャインタフェースを用いた仮想シーン知覚像のスケール再現に向けて. *Visual Computing 2022 予稿集*, pp. 29:1–29:6, 2022.
- [8] 長田 和美, 三輪 智也, 長 篤志, 一川 誠, 水上 嘉樹, 多田村 克己, 三池 秀敏. 知覚される大きさと観察距離の関係を示す拡大率関数: 実空間で得られる視覚印象を表現する画像生成に向けて. *認知科学*, Vol. 15, No. 1, pp. 100–109, 2008.
- [9] Jodie M Plumert, Joseph K Kearney, James F

- Cremer, and Kara Recker. Distance perception in real and virtual environments. *ACM Transactions on Applied Perception*, Vol. 2, No. 3, pp. 216–233, 2005.
- [10] Rebekka S Renner, Boris M Velichkovsky, and Jens R Helmert. The perception of egocentric distances in virtual environments - a review. *ACM Computing Surveys*, Vol. 46, No. 2, pp. 1–40, 2013.
- [11] Etienne Peillard, Thomas Thebaud, Jean-Marie Normand, Ferran Argelaguet, Guillaume Moreau, and Anatole Lécuyer. Virtual objects look farther on the sides: The anisotropy of distance perception in virtual reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces*, pp. 227–236, 2019.
- [12] 小林 秀明, 前原 吾朗. 立体視画像観察時に固視刺激の奥行きが周辺視野の奥行き知覚に及ぼす影響. 基礎心理学研究, Vol. 39, No. 1, pp. 1–9, 2020.
- [13] Xizuo Liu and Guangming Wang. Depth perception optimization based on multi-viewing spaces. In *Proceedings of the 4th International Conference on Computer Science and Application Engineering*, pp. 1–6, 2020.
- [14] Jeffrey S Pierce, Andrew S Forsberg, Matthew J Conway, Seung Hong, Robert C Zeleznik, and Mark R Mine. Image plane interaction techniques in 3D immersive environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, pp. 39–43, 1997.
- [15] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, pp. 99–108, 2017.
- [16] Whie Jung and Woon Tack Woo. Duplication based distance-free freehand virtual object manipulation. In *2017 International Symposium on Ubiquitous Virtual Reality*, pp. 10–13, 2017.
- [17] Ivan Poupyrev, Mark Billingham, Suzanne Weghorst, and Tadao Ichikawa. The go-go interaction technique: non-linear mapping for direct manipulation in VR. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*, pp. 79–80, 1996.
- [18] Laurenz Berger and Katrin Wolf. WIM: Fast locomotion in virtual reality with spatial orientation gain & without motion sickness. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pp. 19–24, 2018.
- [19] 古山 善将, 川畑 博理, 赤羽 亨, 小林 茂, 鈴木 宣也. Ubi-camera: より実用的なジェスチャ型カメラの提案. 情報処理学会インタラクシオン, pp. 653–658, 2012.
- [20] Eureka Works. VRHandsFrame. <https://eurekaworks.jp/software/vrhandsframe/>. 参照: 2022–11–30.
- [21] 菅沼 辰也, 兼松 祥央, 三上 浩司. VR型FPSゲームにおける視線と視野角に応じた敵AIの動的調整に関する研究. 情報処理学会研究報告, Vol. 2019–CG–176, No. 18, pp. 1–4, 2019.
- [22] Shai Avidan and Ariel Shamir. Seam carving for content-aware image resizing. *ACM Transactions on Graphics*, Vol. 26, No. 3, p. 10, 2007.
- [23] 石川 直彦, 朱 臻陽, 茅 曉陽. 知覚サイズを反映した写真生成のための画像自動部分拡大方法. *Visual Computing 2022 発表予稿集*, pp. 42:1–42:5, 2022.

星川 潤



2021年慶應義塾大学理工学部情報工学科卒業。2023年、同大学大学院理工学研究科開放環境科学専攻情報工学専修修士課程を修了（工学）。在学中は、人間の視知覚特性を考慮したVRシステムの研究開発に従事。

藤代 一成（正会員）



1988年理学博士（東京大学）。東京大学、筑波大学、お茶の水女子大学、東北大学を経て、2009年より慶應義塾大学理工学部情報工学科教授。CG・可視化や知的環境メディアに関する研究に従事。第16回CG Japan Award受賞。本会では副会長、評議員、NICOGRAPH International アドバイザリ委員等を歴任。日本工学会、情報処理学会フェロー、画像電子学会名誉会員、ACMシニア会員、IEEEシニア会員およびVisualization Academy会員。現在、日本学術会議連携会員、情報処理学会CGVI研究会主査。