



パストレーシングを用いたプロジェクションマッピングのための 時空間変化を利用したノイズ評価指標

勝部 寛也^{*1} 西田 真也^{*2} 渡辺 義浩^{*1}

Toya Katsube^{*1} and Shin'ya Nishida^{*2} and Yoshihiro Watanabe^{*1}

Abstract – 近年、拡張現実の分野でダイナミックプロジェクションマッピングが注目されている。なかでも、リアリティの高い画像を描画可能なパストレーシングを用いる手法が提案されている。同手法は、誤差が残る描画像を高いフレームレートで投影することで、知覚上で画像の積分を起こし、運動に対してずれのない投影とノイズが少ない知覚像を可能にする。しかし、知覚される視覚ノイズの評価は定性的であった。そこで本稿では、画素値の時間的な変動と空間的な誤差を考慮することで、投影画像の知覚上での積分を利用したプロジェクションマッピングに対するノイズ評価指標を提案する。

Keywords : プロジェクションマッピング, 高速プロジェクタ, 動画評価指標

1 はじめに

近年、拡張現実の研究が盛んである。なかでもプロジェクションマッピングはエンターテインメントなど様々な場面で利用されている [1, 2]。また、従来のプロジェクションマッピングは静止する物体に対してのみに画像を投影するものであったが、近年では運動する物体に対して画像を投影するダイナミックプロジェクションマッピングが提案されている。ダイナミックプロジェクションマッピングでは、非剛体への投影 [3] や、マーカーレスな投影 [4, 5] など様々な手法が考えられてきた。

ダイナミックプロジェクションマッピングの次の課題として、シーンの2次反射や影の再現など、リアリティの向上が挙げられる。この目標に関連して、コンピュータグラフィックスでは、画像の描画手法としてパストレーシングが用いられている。

パストレーシングとは、光線経路をサンプリングすることで、リアリティの高い画像を描画する手法である。光線経路を計算する際、物体表面で光が反射する角度を確率で決めており、シミュレーションによって光源を通る明るい画素値から、光源を通らない暗い画素値まで得られる。最終的な画素値は、各シミュレーションで得られた画素値を計算上で積分することで求められる。

よって、パストレーシングは各画素での光線経路のシミュレーション回数 samples-per-pixel (spp) が多いほど真値に近い値が得られ、少ないほど真値との間に誤差が生まれる。しかし、パストレーシングは光線経

路の計算に時間を要する。したがって、高い spp の画像を描画するには多大な時間を要する欠点がある。

一方、ダイナミックプロジェクションマッピングは、運動する物体に対してずれのない投影をするために、物体のトラッキング、画像の描画、投影までを高速に行う必要がある。そこで、誤差が残る低い spp の画像を高いフレームレートで投影する手法が提案された [6]。同手法は、人が知覚できないほど速く投影画像を切り替えることで、画像が知覚上で積分され、綺麗な投影像が知覚される。また、短時間で描画可能な低い spp の画像を、実対象の運動に整合して高速投影することで、ずれのないマッピング像を知覚させることも可能である。

しかし、同手法を用いても完全に綺麗な知覚像で、ずれのない投影は実現できていない。加えて、知覚される視覚ノイズの評価は定性的なものであった。今後、ダイナミックプロジェクションマッピングのデノイズを考えるうえで、視覚ノイズの量を表す評価指標が必要である。

しかし、実際にプロジェクションマッピングを主観評価するには、毎回被験者を募る必要があり、コストを要する。同様にノイズの評価が必要であるコンピュータグラフィックスにおいて、客観的な画像評価指標として SSIM [7] や FID [8] などがあるが、これらは空間的なノイズ量しか評価できない。また、動画評価指標として、MOVIE [9] などが提案されている。しかし、同指標は 60 fps などの一般的なディスプレイのフレームレートを想定したものであり、約 1000 fps の知覚上の積分を想定したものではない。

一方、高いフレームレートの動画に対してフリッカーを検出する手法が提案されている [10]。また、Mantiukらはコントラスト感度関数をもとにした、時間変化を

^{*1}東京工業大学

^{*2}京都大学

^{*1}Tokyo Institute of Technology

^{*2}Kyoto University

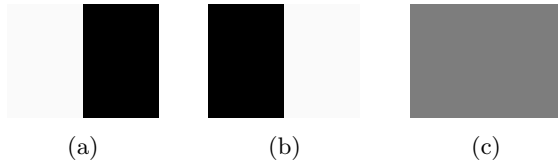


図1 知覚像が一定になる動画のフレーム。(a), (b) は比較動画の奇数, 偶数番目のフレーム, (c) は参照動画全体のフレームである。

Fig.1 Frames of video where the perceptual image becomes constant. (a) and (b) are the odd-numbered and even-numbered frames of the comparison video, and (c) is the all frame of reference video.

考慮した動画指標である FovVideoVDP [11] を提案した。FovVideoVDP は知覚上の積分が起きる高いフレームレートである 240 fps の動画を評価している。しかし、FovVideoVDP は動画指標であり、ダイナミックプロジェクションマッピングのような、約 1000 fps で実世界に投影する動画の評価に利用できるかは不明である。

そこで本稿では、動画評価指標である FovVideoVDP が、プロジェクションマッピングのノイズ評価指標として利用できるか検討する。なお本稿では、運動物体への投影に先立ち、画像の事前描画が可能な静止物体でのノイズ評価を検討した。

2 関連研究

動画の品質は最終的に人間が知覚に沿って判断するものである。しかし、主観評価を得るためには毎回被験者を募る実験が必要になり、コストを要する。そこで、コンピュータグラフィックス業界では MOVIE [9] など様々な動画評価指標が客観評価として提案されてきた。

しかし、動画評価指標は一般的なディスプレイのフレームレートである約 60 fps を想定した指標が多い。知覚上の積分を利用するダイナミックプロジェクションマッピングでは、人が知覚できるフレームレートよりも高い 947 fps のプロジェクタを用いており、知覚上の画像の積分を想定した指標が必要である。

一方、Denes らは時間的に変化する動画に対するフリッカー検出を提案した [10]。同手法は連続するフレーム間の差分をマルチスケール解析し、時空間コントラスト感度関数を用いることでフリッカーを検出する。同指標では、フレームレートが 180 fps で 2 枚の画像が切り替わる動画が評価された。

Mantiuk らは、空間、時間、中心視野を考慮した動画評価指標である FovVideoVDP を提案した [11]。同指標では、動画の時間的な変化の捉え方をモデル化した

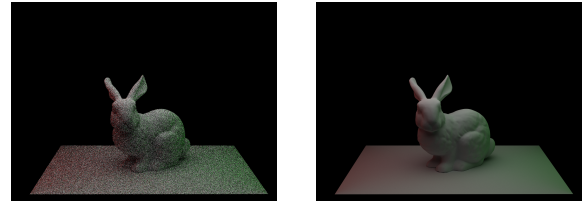


図2 比較, 参照動画に利用したシーン [12]. (a) は 5 spp で描画され, (b) は 10000 spp で描画されている。

Fig.2 Scenes used for comparison and reference videos. (a) is rendered at 5 spp, and (b) is rendered at 10000 spp.

2つの時間チャンネルや、各フレームの空間周波数により、動画を分離する。その後、参照動画と比較動画の差分を計算し、最後にプーリングを行うことで1つの評価値を得る。同指標では、240 fps までの動画が評価された。

3 知覚上の積分を考慮した動画指標の提案

本節では、人の知覚よりも高いフレームレートの動画も対象にしている FovVideoVDP [11] が、知覚上の積分を利用するプロジェクションマッピングに適用できるかを検討する。

知覚上の積分を利用するプロジェクションマッピングでは、人が知覚できる時間よりも高速に画像を投影することで、ノイズが少ない知覚像を提示する。同手法で人が感じるノイズとして、積分された知覚像自体の空間的なノイズの他に、知覚像が時間経過とともに変化する時間的なノイズが考えられる。本節では FovVideoVDP がプロジェクションマッピングに必要な約 1000 fps までにおいても、人の目によって画像が積分される様子を反映しているかを確認する。

最初に、フレームレートが高くなるにつれて、知覚像が一定になる動画を FovVideoVDP を用いて評価する。具体的には、フレームごとの画像は異なるが、人の知覚よりも高いフレームレートで積分された場合、知覚像が一定になる動画を比較動画として用いる。

実際の動画のフレームを図1に示す。参照動画は全てのフレームで各ピクセルの赤、緑、青の値が (125, 125, 125) である動画を用いた。比較動画は奇数フレームでは左半分の値が (250, 250, 250)、右半分の値が (0, 0, 0)、偶数フレームでは左半分の値が (0, 0, 0)、右半分の値が (250, 250, 250) である動画を用いた。

参照動画はいずれのフレームレートにおいても灰色の画像が知覚される。一方、比較動画は低いフレームレートにおいては、画像が積分されずに時間的な変化が見え、高いフレームレートにおいては画像が積分さ

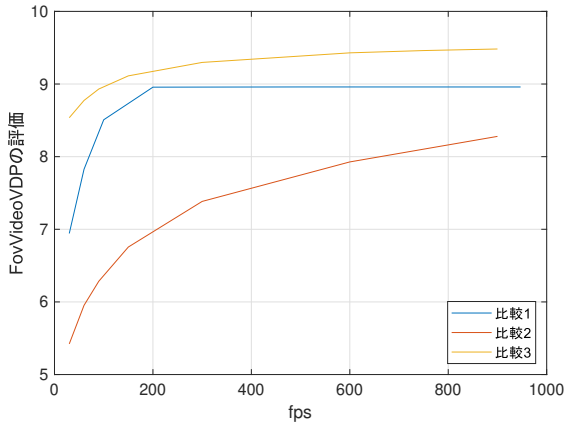


図3 FovVideoVDP の評価結果. 高いフレームレートにおいて, 比較1は知覚像が一定になる場合, 比較2は一定にならない場合であり, 比較3は比較2と同じシーンで知覚像が一定になる場合である.

Fig. 3 Results of FovVideoVDP. At high frame rates, comparison 1 is when the perceptual image is constant, and comparison 2 is when it is not constant, and comparison 3 is the same scene as comparison 2 and the perceptual image is constant.

れ, 灰色の画像が知覚される.

次に, 高いフレームレートでも知覚像が一定ではない動画を FovVideoVDP を用いて比較する. 具体的には, ノイズの無い動画と, 高いフレームレートでもノイズが消えない動画を比較する. 動画はパストレーシングを用いて描画し, 図2のような, 以前我々が収集したプロジェクションマッピングの主観評価 [12] と同じシーンを用いた. 参照動画として 10000 spp で描画した動画を用い, 比較動画として 5 spp で描画した動画を用いた.

パストレーシングではフレームごとにシミュレーション結果が異なる. よって, フレームレートが高くなるほど積分される spp が増え, ノイズが減り続ける. ただし, 今回の 5 spp は 947 fps でも主観評価でノイズが知覚された条件である [12].

最後に, 先に述べた知覚像が一定ではない動画と同じシーンで, 高いフレームレートで知覚像が一定になる動画を用いて比較する. 参照動画として 10000 spp で描画した動画を用い, 比較動画として 80 spp で描画した動画を用いた. 80 spp は 947 fps において, 主観評価でノイズが知覚されなかった 60 spp よりも高い spp である [12].

各比較につき, 複数のフレームレートで評価を行った. 図3に各比較の結果を示す. 比較1は知覚像が高いフレームレートで一定になる場合の評価結果である. 比較2は知覚像が高いフレームレートでも一定にならない場合の評価結果である. 比較3は比較2と同

じシーンであるが, 知覚像が高いフレームレートで一定になる場合の評価結果である.

比較1では 200 fps 付近で動画の評価が一定になっている. これは, 比較動画の知覚像が画像の積分によって一定になった様子を反映している. 一方, 比較2ではフレームレートが高くなるにつれ, 評価が高くなり続けている. これは比較動画に関して, 高いフレームレートでも知覚像が変化し続けるためと考えられる. 比較3では, フレームレートが高くなるにつれ, 評価が徐々に一定になっている. これは, 比較動画のフレームが比較2の動画よりもノイズが少なく, 高いフレームレートで知覚像が一定に近づいているためと考えられる. よって, FovVideoVDP は知覚上の画像の積分によってノイズが知覚できなくなる場合と, 積分後もノイズが残る場合の両方を反映することが確認できた.

4 評価指標パラメータの補正

前節では FovVideoVDP が, 知覚上の画像の積分が起きる高いフレームレートの動画に対して, 知覚像の時間的な変化を反映できることを確認した. 次に, ディスプレイ上の動画指標である FovVideoVDP を, プロジェクションマッピングの評価方法として使用する.

本節では使用にあたり, FovVideoVDP の自由パラメータをプロジェクションマッピングの主観評価をもとに補正した. データとして, 以前収集したプロジェクションマッピングの主観評価のデータ [12] を用いた. 主観評価に用いたシーンはボックスの中に Stanford Bunny [13] を置いたものであり, Stanford Bunny の素材はマット, 粗い金属, 金属の3種類である. 各素材について特定の spp で描画した動画を複数条件用意し, 3素材で合計 36 の spp 条件であった. 主観評価は5段階評価で行い, 数値が高いほどノイズが少ない評価とした.

FovVideoVDP では, 参照動画と比較動画を空間周波数や人の視覚をモデル化した時間チャンネルごとに分離する. よって, 画素, 空間周波数, 時間チャンネル, フレームごとに参照動画と比較動画の誤差が求まる. 同指標では, 上記の誤差をプーリングすることで1つの評価値を計算し, 更に人の主観に沿うように変換する.

FovVideoVDP では誤差を計算する過程, プーリングする過程, 誤差を主観評価に沿うように変換する過程で各々自由パラメータを用いている. 本稿では, 誤差を計算する過程は人の知覚依存であり, 共通であるとして, プーリングと誤差から主観評価への変換の部分を補正する.

FovVideoVDP のプーリングを式 (1), (2) に示す.

$$D_{pooled} = \frac{1}{F^{1/\beta_f}} \left\| w_c \left\| \frac{1}{N_b^{1/\beta_x}} \|D_{b,c}(x)\|_{\beta_x, x} \right\|_{\beta_b, b} \right\|_{\beta_c, c} \right\|_{\beta_f, f} \quad (1)$$

$$\|f(v)\|_{p,v} = \left(\sum_v |f(v)|^p \right)^{1/p} \quad (2)$$

$\beta_x, \beta_b, \beta_c, \beta_f$ は各ピクセル, 各空間周波数, 各時間チャンネル, 各フレームをプーリングする際のノルム距離の自由パラメータである. x はフレーム内のピクセル位置, b は空間周波数, c は時間チャンネル, f はフレームを表す. w_c は各時間チャンネルの重みである.

誤差から評価への変換に関して, プロジェクションマッピングの主観評価データの数値が1から5までのため, 評価の最大値が5になるよう式 (3) で補正した.

$$Q = 5 - \alpha(D_{pooled})^\beta \quad (3)$$

Q は最終的な評価値, α, β は自由パラメータである.

パラメータの補正は, データの56%を用いて行った. 本補正は2段階で行い, 最初にベイズ最適化を行った後, 非線形最適化を行った. Mantiuk らに倣い, 1度に補正するパラメータを少なくするため, プーリング部分の補正と評価への変換部分の補正を分けて行った. 損失関数には, 式 (4) のような FovVideoVDP と主観評価の結果間の RMSE を用いた.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_{s,i} - Q_{o,i})^2} \quad (4)$$

n はデータの数, $Q_{s,i}, Q_{o,i}$ は各データでの主観評価, 評価指標の評価値である.

本補正によって得られた結果を表1に示す. w_s, w_t は式 (1) の w_c に代入される各時間チャンネルの重みである.

表1 FovVideoVDP パラメータの補正結果
Table 1 Calibration result of FovVideoVDP parameters

パラメータ	補正結果
β_x	0.1550
β_b	4.3901
β_c	0.0642
β_f	2.5605
w_s	3.7889
w_t	0.3194
α	0.3160
β	0.3339

5 補正結果の比較と考察

本節では, 補正した FovVideoVDP と他の指標を比較する. 比較に用いるデータは以前収集したプロジェ

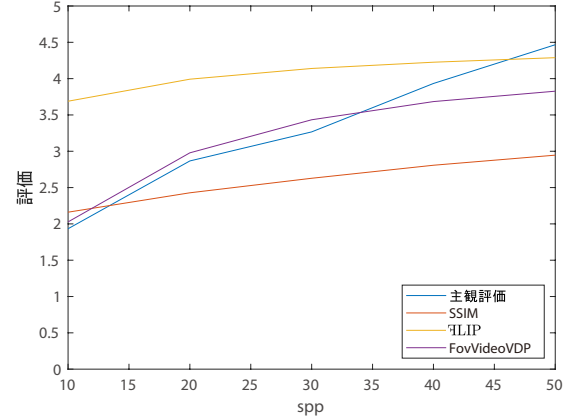


図4 比較用のデータにおける主観評価と客観評価

Fig. 4 Subjective and objective evaluation of comparative data

クションマッピングの主観評価のデータ [12] のうち, 前節で補正に利用しなかったデータを用いた.

比較する指標は, SSIM [7] と FLIP [8] とした. 比較にあたり, 使用するデータの主観評価は1から5の範囲であり, 評価値が高いほどノイズが少ない. 一方, SSIM は0から1の範囲で評価値が高いほどノイズが少なくなる評価, FLIP は0から1の範囲で評価値が高いほどノイズが多くなる評価である. よって, SSIM と FLIP の評価は式 (5), (6) のように補正した.

$$Q_S = 5S \quad (5)$$

$$Q_F = 5(1 - F) \quad (6)$$

Q_S, Q_F は補正後の SSIM, FLIP の評価値, S, F はもとの SSIM, FLIP の評価値である. また, SSIM, FLIP はともに画像評価指標であるが, フレームごとに大きな差が見られなかったため, 特定の1枚のフレームを用いて計算した評価値を動画評価とした.

最初に, 各指標の評価値と主観評価の結果との間で, 式 (4) と同様の RMSE を計算し, 比較する. 各評価指標の RMSE を表2に示す. SSIM や FLIP と比較して, 本稿で補正した FovVideoVDP が最も誤差が少ない結果となった.

表2 評価指標と主観評価の RMSE
Table 2 RMSE between Objective evaluation and Subjective evaluation

評価指標	FovVideoVDP	SSIM	FLIP
RMSE	0.4598	0.7834	0.9242

次に, 各 spp における評価値を比較する. 図4に比較に用いたデータのうち, マットな素材を描画したシーンでの主観評価と評価指標による評価値を示す. 主観評価は各被験者の評価を平均したものであり, 各

指標の評価値は同じ spp での評価を示す。SSIM は単純な輝度値をもとに計算されており、人の知覚との関連性が薄いため、誤差が大きいと考えられる。一方、 Ψ LIP はコントラスト感度などの人の知覚特性を考慮されているが、時間方向のノイズが考慮されていないため、主観評価よりも高い評価値が出ていると考えられる。

6 まとめ

本稿では、動画評価指標である FovVideoVDP がプロジェクションマッピングのノイズ評価指標として利用できるかを検討した。最初に、FovVideoVDP が知覚上の積分を評価できることを確認した。その後、FovVideoVDP をプロジェクションマッピングの評価用に補正した。結果として、SSIM や Ψ LIP などの評価指標と比較して主観評価との誤差が小さい結果が得られた。

今後の課題として、より様々なシーンの描画に対する評価の作成や、運動する物体に対するノイズの評価などが挙げられる。

参考文献

- [1] M. R. Mine, J. Van Baar, A. Grundhofer, D. Rose, and B. Yang, "Projection-based augmented reality in disney theme parks," *Computer*, vol. 45, no. 7, pp. 32–40, 2012.
- [2] A. H. Bermano, M. Billeter, D. Iwai, and A. Grundhöfer, "Makeup lamps: Live augmentation of human faces via projection," in *Computer Graphics Forum*, vol. 36, no. 2. Wiley Online Library, 2017, pp. 311–323.
- [3] G. Narita, Y. Watanabe, and M. Ishikawa, "Dynamic projection mapping onto deforming non-rigid surface using deformable dot cluster marker," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 3, pp. 1235–1248, 2016.
- [4] L. Miyashita, Y. Watanabe, and M. Ishikawa, "Midas projection: Markerless and modelless dynamic projection mapping for material representation," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 1–12, 2018.
- [5] S. Hisaichi, K. Sumino, K. Ueda, H. Kasebe, T. Yamashita, T. Yuasa, U. Lippmann, P. Aswendt, R. Höfling, and Y. Watanabe, "Depth-aware dynamic projection mapping using high-speed rgb and ir projectors," in *SIGGRAPH Asia 2021 Emerging Technologies*, 2021, pp. 1–2.
- [6] T. Nomoto, R. Koishihara, and Y. Watanabe, "Realistic dynamic projection mapping using real-time ray tracing," in *ACM SIGGRAPH 2020 Emerging Technologies*, 2020.
- [7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [8] P. Andersson, J. Nilsson, T. Akenine-Möller, M. Oskarsson, K. Åström, and M. D. Fairchild, "ΨLIP: A difference evaluator for alternating images." *Proc. ACM Comput. Graph. Interact. Tech.*, vol. 3, no. 2, pp. 15–1, 2020.
- [9] K. Seshadrinathan, and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE transactions on image processing*, vol. 19, no. 2, pp. 335–350, 2009.
- [10] G. Denes, and R. K. Mantiuk, "Predicting visible flicker in temporally changing images," *Electronic Imaging*, vol. 2020, no. 11, pp. 233–1, 2020.
- [11] R. K. Mantiuk, G. Denes, A. Chapiro, A. Kaplanyan, G. Rufo, R. Bachy, T. Lian, and A. Patney, "Fovvideovdp: A visible difference predictor for wide field-of-view video," *ACM Transactions on Graphics (TOG)*, vol. 40, no. 4, pp. 1–19, 2021.
- [12] 勝部寛也, 渡辺義浩, "パストレーシングを用いたプロジェクションマッピングのノイズ評価指標に関する検討", 第 27 回バーチャルリアリティ学会大会, 2022.
- [13] "The Stanford 3D Scanning Repository," <http://graphics.stanford.edu/data/3Dscanrep/>.