



人の知覚に寄り添った自然で快適な映像表示 ——人間の視覚情報処理モデルに基づく表示映像の最適化

情報表示技術や表示デバイスの発展に伴い、近い将来、現実空間のあらゆる場所が情報表示に使われるようになるかもしれません。しかし、プロジェクタや透過型ディスプレイを用いた新しい表示技術では、周囲の明るさや背景の模様によって表示画像の見え方が大きく変化するため、映画館のように常に理想的な表示はできません。本稿では、人間の視覚情報処理をモデル化することで、こうした将来の情報表示において、使用場面によらず自然で快適な映像表示を実現するアプローチを紹介します。

キーワード：#メディア表示技術、#人間情報科学、#視覚情報処理モデル

ふきあげ たいき

川上 大樹

NTTコミュニケーション科学基礎研究所

視覚の理解に基づくメディア技術

視覚情報を伝達・共有するための媒体である視覚メディアは、絵画や写真からテレビ、プロジェクタ、スマートフォン、ヘッドマウント型ディスプレイ (HMD) *1 など、多様な形態に発展してきました。現在では、これらのメディアは私たちの生活に欠かせないものとなっています。メディア技術がさらに発展することで、近い将来、あらゆる場所 (空間) に情報提示が行われ、ディスプレイとして利用されると期待されています。では、これらのメディアの上で伝えたい視覚情報を意図どおり伝えるにはどうすればよいでしょうか。現実のシーンをメディア上で再現するには、理想的には再現したい物理空間内の情報をすべて写し取り、それをメディア機器上で再生できることが望めます。しかし、そうした究極のメディア機器はいまだ存在せず、再現の程度はそれぞれの機器が表示できる光の強さや波長、解像度などの物理的な制約に縛られます。こうした物理的制約の中で情報を意図どおり伝えるには、人間が視覚情報をどのように処理し、知覚・認識しているのかを理解することが鍵となります。具体的な例としては、カラーモニタの表示方式が挙げられます。人間の網膜には異なる波長領域の光

(赤・緑・青に対応) に応答する細胞があり、これらの応答の組み合わせで色を知覚する「三色色覚」の仕組みがあります。この知覚を応用し、ディスプレイでは赤・緑・青の3色の光の組み合わせで幅広い色彩を再現しています。また、3DテレビやHMDでは、人間の立体視メカニズムの理解に基づいて3次元の奥行き情報を伝えます。両眼には視点の違いによるわずかな像の差 (両眼視差) が生じ、脳がこの差を処理して奥行きを知覚します。この原理を利用し、左右の目に異なる映像を提示することで、物理的な3D空間を用意することなく、観察者に立体感を体験させることができます。このように、人間の視覚特性を理解し、それを巧みに利用することで、物理世界を完全に再現することなく、効率的に知覚世界を再現することが可能になります。

これまでに挙げた例は、視覚特性に合わせて提示デバイスを設計するというアプローチでした。しかし、現実と仮想の情報が混在する将来の情報提示技術では、表示環境ごとに見え方が常に変化することが想定されます。こうした状況下では、事前のデバイスの設計だけでなく、提示するコンテンツ自体をその場で最適化することが求められます。これを実現するためには、任意の画像に対する知覚を定量的に予測できる「視覚情報処理モデル」を使い、その予測に基づいて提示映像を最適化するアプローチが有効です。

視覚の情報処理モデル

視覚情報処理モデルとは、脳内の視覚情報処理を数理的にモデル化したものです。本稿で扱うモデルの処理の流れを図1に示します。このモデルでは、任意の画像を入力とし、そこから私たちがものを認識する際に用いる特徴を抽出し、その特徴に対する感覚の強さ (知覚量) を予測します。最後に、これらの特徴から、見た目の自然さや見やすさといった映像提示にとって重要な指標を推定します。

では、この「特徴」とは具体的にどのようなものでしょうか。人間は世界の認識や自身の行動のために、網膜に入った情報からさまざまな特徴を抽出して利用しています。特徴抽出の処理は階層的になっており、最初は局所的な領域の色やコントラスト (明暗の差) といった単純な特徴を抽出し、その後はこれらの特徴を統合しながら、方位や形状、テクスチャ、さらには顔や物体、風景など、複雑で大域的な特徴の検出へと進んでいきます。しかし、こうした特徴抽出処理のうち、具体的に実用可能な計算モデルとして確立されている部分は限られています。以下では、今回取り上げる研究例での使用実績のある低次の視覚情報処理に焦点を当てて解説します。

低次の視覚情報処理モデルによる特徴抽出の具体的な処理を図1の破線の枠内に示しました。まず、「色・輝度分解」について説明します。私たちの網膜には赤・緑・青の3つの波長帯に対応した錐体というセ

*1 ヘッドマウント型ディスプレイ：頭部に装着して使用するディスプレイ装置。映像を直接目の前に映し出すことで、没入感の高い視覚体験が得られます。

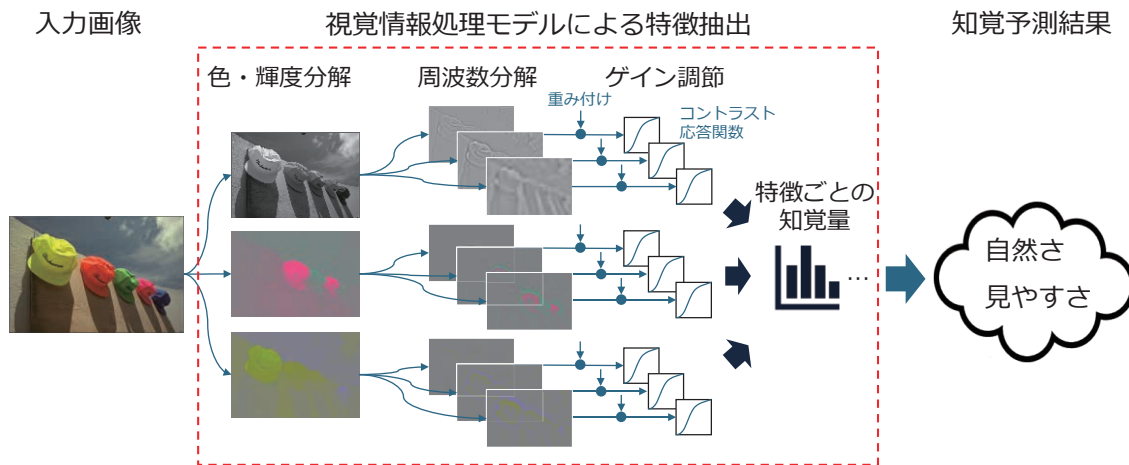


図1 視覚情報処理モデルの処理の流れ

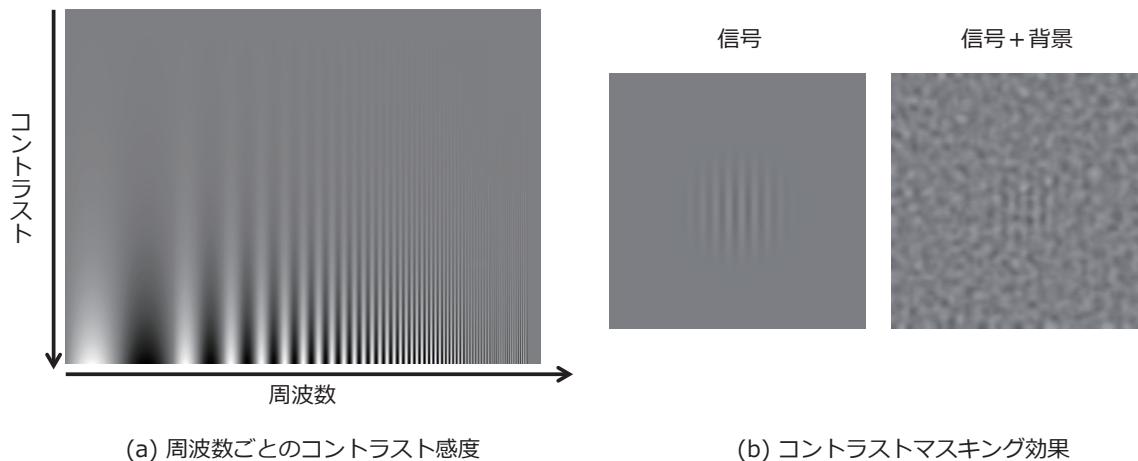


図2 コントラスト知覚のデモ

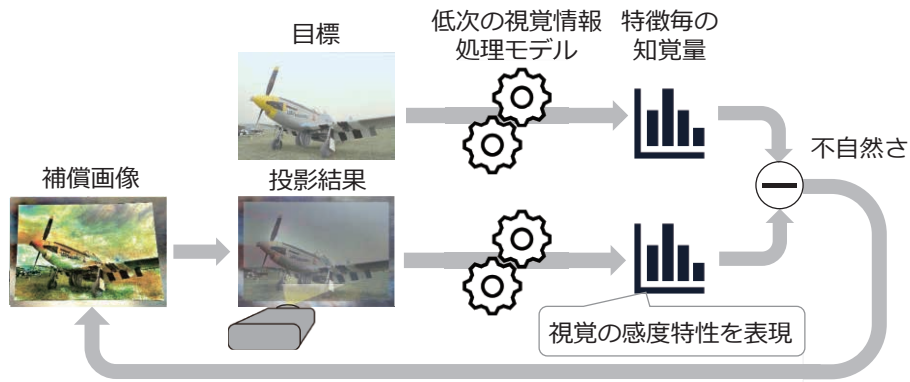
ンサがあります。これらのセンサに入力された光の情報は、色の違いを際立たせつつ、効率的に色情報を伝達するため、反対色と呼ばれる色の差分という形式に変換されて、後の処理に進みます。「色・輝度分解」処理はこの色の処理メカニズムを模倣しており、入力画像の赤・緑・青の各色チャンネルを足し引きすることで、明るさを表現する輝度の成分と、赤と緑、青と黄の色の差をそれぞれ表現する2つの反対色成分とに画像を分解します。

次に、各色成分に対応する画像はそれぞれ周波数分解処理を受けます。ここでの周波数は、パターンの空間的な細かさを表す値です。人間の視覚系には、さまざまな細かさを選択的に応答する神経細胞があり、これらの応答が網膜像の中の周波数特徴を表現しています。視覚情報処理モデルでは、

この周波数についての情報表現を再現するために「畳み込み」という画像処理が用いられます。畳み込みによって、さまざまな大きさの周波数ごとのコントラストを表す画像が得られます。さらに、最後に各周波数成分に重みをかけることで、人間の視覚系が異なる周波数に対して持つ感度の違い⁽¹⁾を反映します。図2(a)に、こうした周波数ごとの感度の違いを体感できるデモを示しました。この画像では、右側に行くほど周波数が大きく（パターンが細かく）なり、上側に行くほど物理的なコントラストが低くなっています。ここで、物理的なコントラストは周波数間で一定であるにもかかわらず、縞模様が見える部分と見えない部分の境界が上に凸のカーブを描いて見えます。このカーブは周波数ごとの視覚系の感度の違いを表しています。すなわち、視覚系は

中間の細かさに対してもっとも感度が高く、非常に粗いパターンや非常に細かいパターンには感度が低いという特性を持っているのです。

最後に、「ゲイン調節」について説明します。ゲイン調節はコントラストの知覚的な強さと密接にかかわっています。視覚系は広範囲のコントラストに対応できるように神経応答のゲインを調節しており、物理的なコントラストの増加に対して、はじめは急激に応答が増加しますが、高コントラスト域では次第に緩やかになります⁽²⁾。この様子を、横軸を物理コントラスト、縦軸を神経応答として表したのが、図1中に示したコントラスト応答関数です。コントラストマスキング効果は、このゲイン調節の分かりやすい例です。図2(b)の左右の画像には同じコントラストで縞模様が埋め込ま



(a) 視覚情報処理モデルに基づく補償画像の最適化



(b) 投影結果の比較

図3 実物体表面の自然な見た目操作

れていますが、右側の背景ノイズ上にある縞模様は視認性が大きく低下して見えます。これは、背景ノイズによってすでに神経細胞が強く応答しており、縞模様による追加の応答が相対的に小さくなることから説明できます。視覚情報処理モデルでは、こうしたゲイン調節メカニズムを数理的に表現することで、各特徴に対する視覚系の知覚量を定量的に予測します。

視覚情報処理モデルをつかった表示映像の最適化

以上のように、視覚情報処理モデルは、任意の画像を視覚系の感度特性を反映した特徴量に変換します。ここからは、この視覚情報処理モデルを使うことでどういったことが実現できるのかについて、私たちが行ってきた研究事例を基に説明します。

■実物体表面の自然な見た目操作

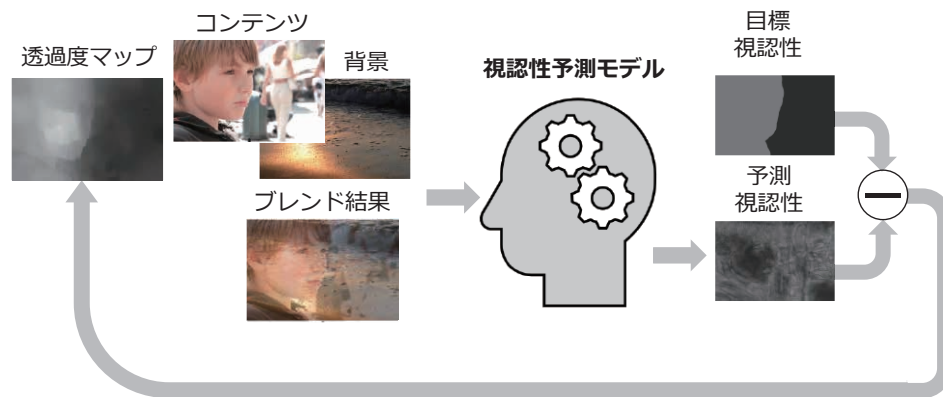
まず取り上げるのは、プロジェクタを使った空間拡張現実に関する研究です。この技術はプロジェクションマッピングとも呼ば

れ、投影対象の実物体表面の見た目を操作することができます。今はまだ大規模なショーやデモンストレーションで用いられることがほとんどですが、将来的にはより身近なところでさまざまな情報提示に用いられる可能性があります。そのような場面で特に解決が求められる技術的課題の1つが、投影対象自体が持つ模様と、投影した画像とが混ざり合ってしまうという問題です。

この解決手段として、色補償と呼ばれる技術があります。この技術では、カメラで投影面を撮影し、その情報を基に投影面の模様を打ち消すように投影画像を修正します⁽³⁾。プロジェクタから光を消滅させるマイナスの光を出力することはできないため、例えば投影面が赤い模様を持つ場合、シアンの色を投射して色味を打ち消し、そこからさらに提示したい色を足すことで投影画像をつくります。しかし、明るい環境光の下では、投影面自体の模様が持つコントラストが大きくなるため、打ち消すのに必要な光の強さも非常に大きくなります。一般

的なプロジェクタではそこまで強い光を出力できず、模様を補償しきれない場合も出てきます。

こうした問題を解決するのに、視覚の感度特性をうまく利用することが非常に有効です。人間の感度が低い特徴を多少犠牲にしながら、感度が高い特徴を優先して再現することで、物理的には正しく補償できていなくても、知覚的には自然に見えるような結果を得ることができるためです。私たちの研究では、これを実現するために、低次の視覚情報処理モデルを用いました⁽⁴⁾。具体的な手順を図3(a)に示します。まず、目標とする画像と投影結果を撮影した画像とをモデルに入力して知覚の特徴表現に変換します。この特徴量には、視覚の感度特性が表現されているため、これらの特徴間の差の大きさを投影結果の知覚的な不自然さとして定義できます。そして、この不自然さができるだけ小さくなるように補償画像を最適化します。これにより、物理的には目標と一致しないものの、知覚的には自然な投影結果を自動的に得ることができま



(a) 視覚情報処理モデルを使った視認性予測による透過度の最適化



(b) ブレンド結果の比較

図4 視認性予測に基づく「見やすい」透過表示

す。実際の最適化結果の例を図3(b)に示します。物理ベースの手法では投影面の模様がほとんど補償しきれていないのに対し、視覚情報処理モデルを使った知覚ベースの補償結果は、見た目としては目標画像にかなり近い結果が得られています。

これと同様の方法は、私たちが以前に開発したプロジェクション技術である「変幻灯」の課題解決にも用いられました。変幻灯は、物体の動きを表現する白黒の動画パターンを投影することで、静止した実物体があたかも動いているかのような錯覚を生み出す技術です⁽⁵⁾。しかし、自然に見える動きの大きさには限界があり、従来は手動で細かな調整が必要でした。この課題を解決するため、視覚情報処理モデルを活用し、投影結果の自然さを予測して動き情報を自動的に最適化する手法を開発しました⁽⁶⁾。これにより、不自然さを感じない範囲で最

大の動きを得られるようになり、ユーザの表情に合わせて絵画の表情を動かすなど、インタラクティブなアプリケーションでも変幻灯を効果的に活用できるようになりました。

■現実シーンへの「見やすい」透過表示

VR (Virtual Reality)^{*2} や AR (Augmented Reality)^{*3} といった視野全体を覆うことが想定されるメディア技術では、表示する情報が視界を遮らないように半透明で表示することがよくあります。しかし、背景となる現実シーンが常に変化する状況下では、その上に重畳したコンテンツの視認性を一定に保つのは一般に困難です。これは、前述したコントラストマスキング効果の例から分かるように、視認性が背景のコントラストに大きく影響を受けるためです。しかし、視覚情報処理モデルを使えば、こうした半透明画像の視認性の変化も定量

的に予測することが可能になります。私たちの研究では、この視覚情報処理モデルに基づく視認性予測モデルを用いることで、透過度を自動で調節する技術を提案しました⁽⁷⁾。図4(a)に示したとおり、この手法では、ユーザが物理的な透過度ではなく、目標とする視認性を指定します。そして、コンテンツと背景が与えられたときに、最適化したブレンド画像に対して視認性予測モデルが視認性を予測します。その後、目標視認性と予測視認性の差分をとって、この差分が小さくなるように透過度マップを最適化します。図4(b)に結果の例を示します。

*2 VR：仮想現実感とも呼ばれます。コンピュータを用いて作り出された仮想の映像世界にユーザを没入させる技術。

*3 AR：拡張現実感とも呼ばれます。現実の世界に仮想情報を重ね合わせて表示することで、直感的で便利な情報提供を可能にします。

ここでは、2つの背景に同じコンテンツを透過表示した結果を示しています。通常の透過ブレンドの結果では、同じ透過度を設定していても、コンテンツ画像の視認性が背景によって大きく異なることが分かります。一方、提案手法では、同じ視認性マップを目標として透過度を最適化してからブレンドします。その結果、異なる背景でも一貫した視認性でコンテンツを表示できます。このように、提案手法では、ユーザが知覚量である視認性を操作することで、より直感的に意図したような透過合成ができるようになります。こうした方法を用いることで、VRやARなどのインタラクティブメディア上で常に快適な視認性を保つ半透明表示などといった活用も期待できます。

今後の課題と展望

視覚情報処理モデルを用いたコンテンツの最適化は今後のメディア技術において、ますます重要性が増していくと考えられます。しかし、このアプローチにはまだ取り組むべき課題が多く残されています。まず、現在私たちが映像の最適化に活用できている視覚情報処理モデルは、人間の脳内で行われている複雑な視覚情報処理のうち、入口付近に相当するごく一部に過ぎません。今後の研究の発展のためには、中～高次の情報処理のモデル化を進めていく必要があります。具体的には、テクスチャ・奥行き・運動・質感などを予測できるようにすることで、これらの印象を変えずに提示映像をより柔軟に適応させることが可能になるでしょう。

ただし、高次なモデル構築には、低次の視覚モデル構築で用いられてきたような、視覚情報処理を小さなサブプロセスごとに理解して組み上げていく要素構成的なアプローチでは限界があります。そこで、今後は深層学習モデルの活用が鍵となります。深層学習モデルは、物体認識などのタスクを設定することで、入力画像からタスク回答までの複雑な情報処理を自動的に学習します。また、物体認識用に訓練して得られた深層学習モデルと人間の脳の情報処理の類似性もさまざまな観点から明らかになっている⁽⁸⁾という点も注目すべき知見です。

しかし、これらのモデルが人の知覚と定量的に一致するわけではなく、そのままでは映像最適化には使えないという課題があります。また、性能が向上するにつれて、人間の知覚との乖離が大きくなってきているという報告もあります⁽⁹⁾。今後は、人間の知覚との整合性を高めつつ、深層学習モデルを訓練する方法を解明する必要があります。

さらに、視覚情報処理のモデル化を進めるとともに、人間にとっての自然さや快適さの必要条件を明らかにしていくことも重要です。エッシャーの無限階段の例にみられるように、人間は物理的にはあり得ない状況でも、一見すると自然に感じてしまうことがあります。したがって、人間が自然に感じられる映像の分布は、物理に忠実に再現された映像分布よりも広大な裾野をもって広がっていると考えられます。こうした分布の広がりを正しく推定できるようにしていくことで、さまざまな環境・物理的制約の中で映像表現の幅をさらに拡張できると期待されます。

本稿で紹介した成果の一部は東京大学との共同研究によるものです。

参考文献

- (1) F. W. Campbell and J. G. Robson: "Application of Fourier Analysis to the Visibility of Gratings," *The Journal of Physiology*, Vol. 197, No. 3, pp. 1551-1566, 1968.
- (2) D. J. Heeger: "Normalization of Cell Responses in Cat Striate Cortex," *Visual Neuroscience*, Vol. 9, No. 2, pp. 181-197, 1992.
- (3) M. D. Grossberg, H. Peri, S. K. Nayar, and P. N. Belhumeur: "Making One Object Look like Another: Controlling Appearance Using a Projector-Camera System," *Proc. of CVPR 2004*, Jun 2004.
- (4) R. Akiyama, T. Fukiage, and S. Nishida: "Perceptually-based optimization for radiometric projector compensation," *Proc. of IEEE VR 2022*, March 2022.
- (5) T. Kawabe, T. Fukiage, M. Sawayama, and S. Nishida: "Deformation Lamps: A Projection Technique to Make Static Objects Perceptually Dynamic," *ACM Transactions on Applied Perception*, Vol. 13, No. 2, pp. 1-17, 2016.
- (6) T. Fukiage, T. Kawabe, and S. Nishida: "Perceptually based adaptive motion retargeting to animate real objects by light projection," *IEEE Transaction on*

Visualization and Computer Graphics, Vol. 25, No. 5, pp. 2061-2071, 2019.

- (7) T. Fukiage and T. Oishi: "A content-adaptive visibility predictor for perceptually optimized image blending," *ACM Transaction on Applied Perception*, Vol. 20, No. 3, pp. 1-29, 2023.
- (8) D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo: "Performance-optimized hierarchical models predict neural responses in higher visual cortex," *PNAS*, Vol. 111, No. 23, pp. 8619-8624, 2014.
- (9) M. Kumar, N. Houlsby, N. Kalchbrenner, and E. D. Cubuk: "Do better ImageNet classifiers assess perceptual similarity better?," *TMLR*, Sept 2022.



吹上 大樹

NTT研究所では、今後も視覚情報処理の科学的理解に取り組みながら、人間の特性を活かした新しいメディア技術の実現に貢献していきます。

◆問い合わせ先

NTTコミュニケーション科学基礎研究所
企画部 情報戦略担当
TEL 0774-93-5020
FAX 0774-93-5026
E-mail cs-jousen-ml@ntt.com