

ビデオマグニフィケーションが ユーザーの映像変化知覚と注視領域に与える影響の分析

武田 翔一郎^{1,a)} 亀田 明男¹ 磯貝 愛¹ 木全 英明¹

概要: ビデオマグニフィケーションは映像中に潜んでいるヒトの視覚では知覚できないほど微細な変化を強調し可視化する映像処理・合成技術である。これまでに、映像中の微細な変化のみを選択的に検出・強調・可視化するために、様々な手法が提案され、ビデオマグニフィケーションの適用範囲は大きく広がっている。しかし現在までに、ビデオマグニフィケーションを適用することで、ユーザーが微細な変化に気が付くようになるのか（これを映像変化知覚と定義する）、また、その際のユーザーの注視領域は変化しているのか、といったビデオマグニフィケーションがユーザーに与える影響の分析は行われていない。そこで本稿では、様々な強調率を用いたビデオマグニフィケーションをいくつかの映像サンプルに適用し、その前後においてユーザーの映像変化知覚や注視領域がどのように変化するかを主観評価実験を通じて明らかにする。

キーワード: ビデオマグニフィケーション, 映像変化知覚, 注視領域

An Analysis of Video Magnification Effect on Image Change Perception and Gaze Region

Abstract: Video magnification reveals subtle changes invisible to the naked eye. Recently, many video magnification methods have been proposed to pass only subtle change under various disturbance in video. However, it has been not clear if users can perceive subtle image changes and user's gaze region is changed after applying video magnification method. In this paper, we evaluate user's image changes perception and gaze region before and after applying video magnification method with various amplification factor.

Keywords: Video magnification, image changes perception, gaze region

1. はじめに

映像中に存在する微細な変化を強調して表現する映像合成技術であるビデオマグニフィケーションは、今までヒトの視覚では捉えることが困難であった、スポーツ中の肌や筋の微細な変形や、エンジンなどで生じる異常振動、または、音を奏でる弦楽器の振動など、様々な現象を容易に捉えることができるようになる映像合成技術として、近年注目を浴びている [1][2][3][4][5][6][7]。しかし、微細な変化だけでなく、被写体やカメラが横に移動したり、飛び跳ねたりするような、ゆっくり又は素早く大きい運動が映像中に

存在する場合、微細な変化だけを検出することが困難になり、ノイズが発生することが知られている。

そこで近年は、ゆっくりで大きい運動が存在する映像から、微細な変化だけを検出・強調することが可能ないくつかの手法 [4][5][6] が提案され、さらにそれらを発展させた武田らの研究 [7] によって、素早く大きい運動が映像中に存在している場合でも微細な変化のみを選択的に検出・強調することが可能な手法が提案され、ビデオマグニフィケーションの適用範囲は大きく広がっている。

しかし現在までに、ビデオマグニフィケーションを適用することで、ユーザーが微細な変化に気が付くようになるのか（これを映像変化知覚と定義する）、また、その際のユーザーの注視領域は変化しているのか、といったビデオマグニフィケーションがユーザーに与える影響の分析は行

¹ 日本電信電話株式会社
メディアインテリジェンス研究所 高臨場映像通信技術グループ
Hikarinooka1-1, Kanagawa, 239-0847, Japan

^{a)} shoichiro.takeda.us@hco.ntt.co.jp

われていない。特に、ビデオマグニフィケーションの強調率と、ユーザーの映像変化知覚や注視領域の関係性も未だ未解明である。

そこで本稿では、様々な強調率を用いたビデオマグニフィケーションをいくつかの映像サンプルに適用し、その前後においてユーザーの映像変化知覚や注視領域がどのように変化するかを主観評価実験を通じて明らかにする。

2. 関連研究

ビデオマグニフィケーションは、映像中のピクセル毎の色、輝度、又は位相変化（位相は映像中の局所的な運動情報を反映している [8]）を捉えることで、被写体の微細な色、輝度、又は運動を検出・強調する手法である [1][2][3][4][5][6][7]。先行研究で一般的に用いられている手法の流れとしては、まず、映像の各フレーム画像に対して画像ピラミッドを構築し多重解像度表現を取得することで、映像のピクセル毎に様々な大きさの色、輝度、又は位相変化を取得する。その後、得られた各種変化に対して、任意の応答周波数を持つ時系列フィルタ（例えば、バンドパスフィルタなど）を適用し、任意周波数の微細変化を得る。そして、取得した微細変化を α 倍して元の各種変化に加算し、最後に画像ピラミッドを再構成することで、映像中の微細な変化のみが強調された合成映像を得ることができる。しかし、従来手法では、映像中の微細な変化と被写体の大きな運動をピクセル内で切り分けることができず、被写体が大きく動く場合、ノイズが発生するという課題を抱えていた。

この課題、つまり、大きな運動が存在する中でも微細変化のみを検出・強調するために、レイヤーに基づいた手法が提案されている [4][5]。これらの手法は、ユーザーによる強調領域の指定 [4]、または、デプスカメラを用いて強調領域と背景を切り分けることで [5]、大きい運動による影響を軽減する手法である。しかしながら、これらの手法は、人手による操作や [4]、デプスカメラなどの特殊な撮影環境が要求され [5]、手法の適用範囲が限定的である。

一方で Zhang らは、ゆっくりで大きい運動は時間スケールで見ると線形近似できることに着目し、微細な加速度変化のみを検出・強調することを提案した [6]。この手法は、ゆっくりで大きい運動が映像中に存在していても、追加の手間や情報なしで微細変化のみを強調することを可能にした。しかし、映像中に素早く大きい運動が存在すると、それを誤認識し、大きなノイズを生じさせてしまうという課題を持っている。

そこで近年、武田らによって素早く大きい運動が存在する中でも微細な変化のみを検出・強調する新しい手法が提案された [7]。この手法は、素早く大きい運動と比較して微細な変化は滑らかに変化するという観察結果を基に、滑らかさを評価する指標として神経科学や機械工学の分野で用いられてきた躍度 (jerk) という指標に着目し、新しい時

空間フィルタである jerk-aware filter を設計している。この jerk-aware filter を Zhang らの手法に適用することで、ゆっくり又は素早く大きい運動が映像中に存在する場合でも微細な変化のみを選択的に検出・強調することが可能になっている。

3. Jerk-Aware Video Motion Magnification [7]

本稿では、武田らが提案した素早く大きい運動にも頑健性の高いビデオマグニフィケーション手法 [7] を評価に用いる。本稿では、微細な“運動”の強調に関して主観評価実験をおこなうため、ここでは特に微細な運動変化を強調する手法である、Jerk-Aware Video Motion Magnification [7] の詳細について述べていく。

入力映像を $I(x, y, t)$ とする時 ((x, y) はピクセル位置、 t は時間)、ある解像度を持つピラミッドの階層を l 、空間周波数を ω 、及び空間方向選択性を θ とした時、複数次解像度において運動情報を表す位相変化 $\phi_{\omega, \theta}^l(x, y, t)$ を得るため、complex-steerable filter $\psi_{\omega, \theta}^l$ を入力映像に対して以下のように適用する：

$$\psi_{\omega, \theta}^l \otimes I(x, y, t) = A_{\omega, \theta}^l(x, y, t) e^{i\phi_{\omega, \theta}^l(x, y, t)}. \quad (1)$$

ここで、 $A_{\omega, \theta}^l(x, y, t)$ は振幅変化を表している。

その後、微細な位相変化検出するために、Zhang らの手法 [6] と同じように任意周波数 f に応答性を持たせたガウシアンフィルタ $G_f(t)$ を時系列方向に畳み込んだ上で、2回微分 $\frac{\partial^2}{\partial t^2}$ を適用し、微細な位相変化 $C_{f, \omega, \theta}^l(x, y, t)$ を以下のように検出する：

$$C_{f, \omega, \theta}^l(x, y, t) = \frac{\partial^2 G_f(t)}{\partial t^2} \otimes \phi_{\omega, \theta}^l(x, y, t). \quad (2)$$

ここで検出された微細な位相変化 $C_{f, \omega, \theta}^l(x, y, t)$ は、位相変化の中で線形近似可能なゆっくりで大きい運動を無視しているが、素早く大きい運動をご検出し、混入している。Jerk-Aware Video Motion Magnification [7] では、微細変化は素早く大きい運動と比較して時系列的に滑らかに変化するという観察結果（図 1）から、時系列データの滑らかさを評価する指標として神経科学や機械工学の分野で用いられてきた躍度と呼ばれる指標を利用し、微細な変化のみを透過させ、素早く大きい運動の影響を除去するフィルタである jerk-aware filter を設計する。

まず、式 (1) で得られた位相変化 $\phi_{\omega, \theta}^l(x, y, t)$ に対する躍度 $J_{f, \omega, \theta}^l(x, y, t)$ を計算するために、任意周波数 f に応答性を持たせたガウシアンフィルタ $G_f(t)$ を時系列方向に畳み込んだ上で、躍度を導出するための三回微分 $\frac{\partial^3}{\partial t^3}$ を以下のように適用する：

$$J_{f, \omega, \theta}^l(x, y, t) = \frac{\partial^3 G_f(t)}{\partial t^3} \otimes \phi_{\omega, \theta}^l(x, y, t). \quad (3)$$

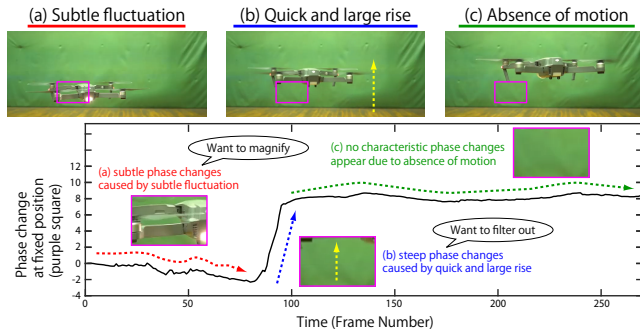


図 1: 映像中の運動 (位相) 変化. ドローンの微細な揺れによる微細な位相変化と, 素早い上昇による急峻な位相変化が生じている. この急峻な変化と比較して, 微細な変化はその振幅の小ささゆえに, 滑らかに変化していると観察できる.

次に, この位相変化から算出された躍度を基に, 位相変化の滑らかさを求める. 滑らかな位相変化の時には 1, 素早く大きい位相変化が生じている時には 0 になるように, 以下のように躍度の値を正規化し, 反転させる.

$$nJ = \frac{|J| - \min |J|}{\max |J| - \min |J|}, \quad (4)$$

$$\text{smoothness}_{f,\omega,\theta}^l(x, y, t) = 1 - nJ. \quad (5)$$

ただしここで, $J := J_{f,\omega,\theta}^l(x, y, t)$, $nJ = nJ_{f,\omega,\theta}^l(x, y, t)$ とした.

最後に, フィルタの抑制効果を調整可能にするために, ハイパーパラメーター β を導入し, jerk-aware filter $JAF_{f,\omega,\theta}^l(x, y, t)$ を得る:

$$JAF_{f,\omega,\theta}^l(x, y, t) = \text{smoothness}_{f,\omega,\theta}^l(x, y, t)^\beta. \quad (6)$$

また, 上記のフィルタは, 映像ピラミッドの各階層 l 毎に求められるが, 階層毎の解像度の違いにより, 算出される躍度の大きさが変わってくる[9]. そこで, ダウンサンプルファクタ ($0 < \lambda < 1$) を用いて, 階層毎に求められた jerk-aware filter を以下の式で補正する:

$$JAF_{\lambda,f,\omega,\theta}^l(x, y, t) = JAF_{f,\omega,\theta}^l(x, y, t)^{1/\lambda}. \quad (7)$$

さらに, jerk-aware filter は, オプティカルフロー等で用いられてきた疎密戦略と同じように, 階層を跨いだ情報のやり取りによって補正する必要がある[9]. したがって, 階層 l のフィルタをそれよりも上の階層 $l+N$ のフィルタで補正する, 伝搬補正というものを以下の式で実行する:

$$pJAF_{\lambda,f,\omega,\theta}^l(x, y, t) = \prod_{i=l}^{l+N} \text{res}(JAF_{\lambda,f,\omega,\theta}^i(x, y, t), l). \quad (8)$$

$\text{res}(JAF_{\lambda,f,\omega,\theta}^i(x, y, t), l)$ は, バイキュービック補間を示している. これらの補正を通じて, 補正された jerk-aware filter である $pJAF_{\lambda,f,\omega,\theta}^l(x, y, t)$ を得ることができる.

そして, 上記のフィルタを式 (2) に対して以下のように適用することで, ゆっくり又は素早く大きい運動を無視し

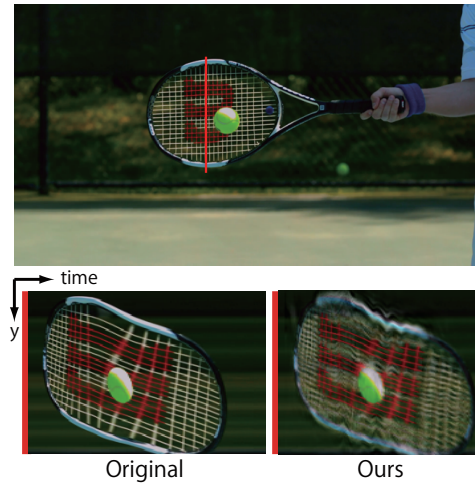


図 2: Tennis: 素早く大きい運動と微細な変化. 下段 (左) は, 元映像において赤線のスライス時間を時間方向にスタックした画像. 下段 (右) は, 強調後の映像において赤線のスライス時間を時間方向にスタックしたものであり, ボールが当たった瞬間にガットに広がる微細な変化が強調されている.

表 1: 評価コンテンツ

Video	Resolution	time(s)
Golf	720 × 480	12.0
Gun	720 × 576	11.5
Tennis	640 × 360	12.07
Wood	640 × 360	12.13
Slam dunk	582 × 328	15.0

ながら, 微細な位相変化 $\hat{C}_{f,\omega,\theta}^l(x, y, t)$ のみを選択的に検出する:

$$\hat{C}_{\lambda,f,\omega,\theta}^l(x, y, t) = pJAF_{\lambda,f,\omega,\theta}^l(x, y, t) \circ C_{f,\omega,\theta}^l(x, y, t). \quad (9)$$

ここで \circ は要素積を表している.

最終的に, 強調率 α を乗算し, 元の位相変化に加算することで, 微細な運動変化のみが強調された位相変化 $\hat{\phi}_{f,\omega,\theta}^l(x, y, t)$ を以下のように得ることができる:

$$\hat{\phi}_{f,\omega,\theta}^l(x, y, t) = \phi_{f,\omega,\theta}^l(x, y, t) + \alpha \cdot \hat{C}_{f,\omega,\theta}^l(x, y, t). \quad (10)$$

4. 主観評価実験

本章では, 主観評価で用いた評価コンテンツおよび評価・分析手法に関して説明した後, 実験方法について説明する.

4.1 評価コンテンツ

素早く大きい運動に微細な変化が含まれる 4 種類のコンテンツを用意した (表 1). 例えば, 評価コンテンツ Tennis には, スイングという素早く大きい運動中に, ボールがラケットに当たった瞬間にガットに伝わる微細な変化が含まれている (図 2).

各コンテンツに対して, 強調率 α は $\alpha = [6, 12, 18, 24, 30]$

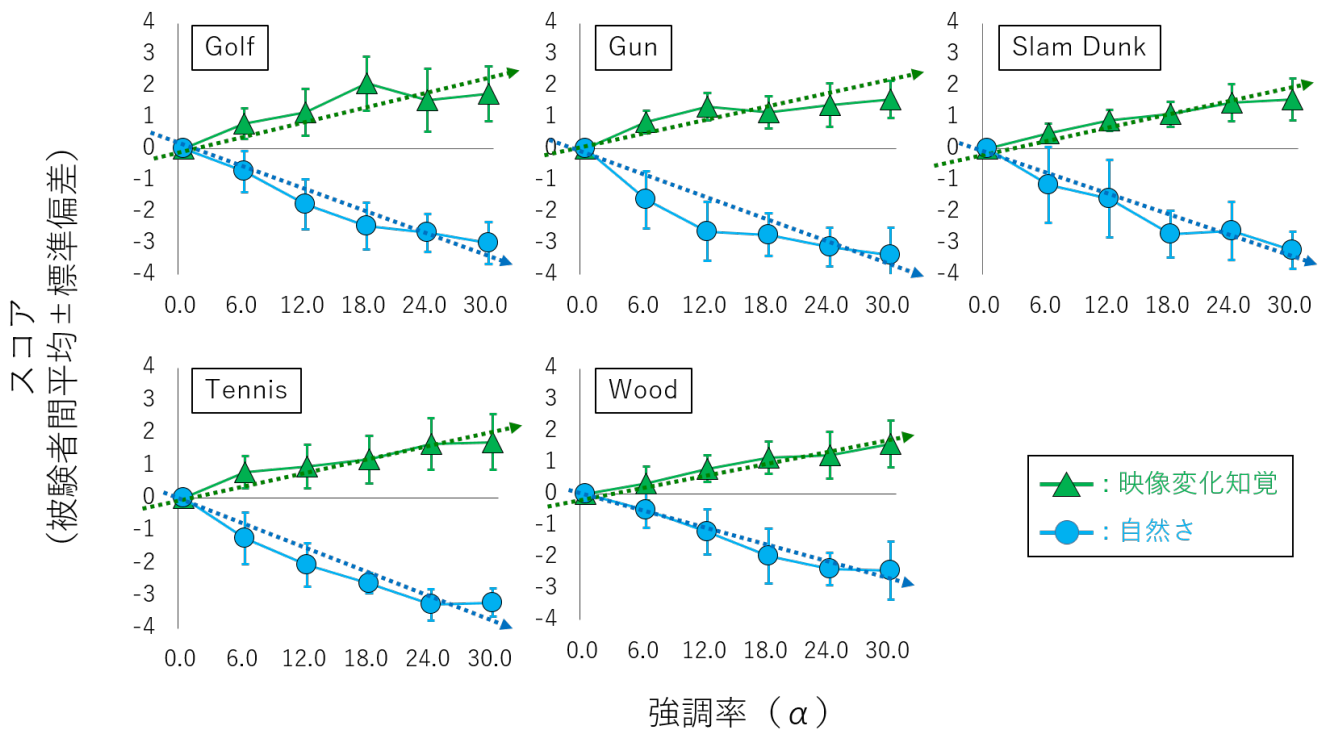


図 3: ビデオマグニフィケーションの強調率を変化させたときの、コンテンツごとのユーザーの映像変化知覚及び映像に対する自然さの推移。全ての場合で線形回帰係数は有意な値を示している ($p < 0.05$).

と変化させ、微細な変化を強調した映像を作成した。以降、微細な変化が強調されていない元映像は、強調率 $\alpha = 0$ の映像として表現する。

4.2 実験方法

微細な変化の強調率の大きさによって、ユーザーの映像変化知覚と注視領域がどのように変化するかを評価するため、被験者に基準映像(元映像)と強調率 α を変化させた評価映像を対にして 2 回提示し、2 回目の提示時に両映像に対して下記で述べるそれぞれの評価を行うよう指示した。

このとき、DCR (Degradation Category Rating) 法 [10] を参考に、基準映像の後に評価映像を提示することで順序効果の影響を抑制している。被験者は 20 代から 50 代の男女 12 名で構成した。

4.3 映像変化知覚の評価と分析方法

微細な変化の強調率 α の増加に伴い、ユーザーの映像変化知覚がどのように変化するかを明らかにするため、下記のような評価観点を設定した。

- 変化の気づきやすさ(映像変化知覚) : 微細な変化を強調することで、微細な変化に気づきやすくなったか、映像中の「変化の度合い」を 5 点: 強い~1 点: 弱いの 5 段階で評価
- 自然さの維持効果 : 微細な変化を強調することで、映像が不自然であると感じるようになるか、映像の

「自然さ」を 5 点: 自然である~1 点: 自然ではないの 5 段階で評価

被験者は、上記評価観点の評価尺度が連続的に表現されているスケール上に、棒線で評価を記入した。

実験後、記入された棒線の位置を小数第一位まで測り、それを被験者の各映像の評価値として換算した。そして、基準映像に比べてどのくらいの差があるか評価するため、DSCQS(Double Stimulus Continuous Quality Scale) 法 [11] を参考に、微細な変化を強調した評価映像の評価値と基準映像の評価値との差分を映像変化知覚を評価する最終的なスコアとした。その後、強調率と最終スコアの関係性を線形回帰を用いて定量的に分析した。

4.4 注視領域の評価と分析方法

微細な変化の強調率 α の増加に伴い、ユーザーの注視領域がどのように変化するかを明らかにするため、被験者は注視した映像領域を用意された画像(映像中のある 1 フレーム)上に書き込み、自身の映像注視領域を評価した。

実験後、その映像領域の変化度合いを定性的に分析した。

5. 分析結果

5.1 ユーザーの映像変化知覚の変化

図 3 に、スコアの平均を縦軸、強調率 α を横軸として、ユーザーの映像変化知覚、並びに不自然さに関する主観評価実験の結果を示す。

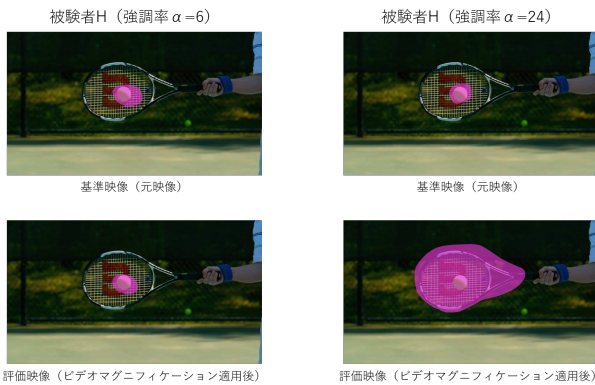


図 4: コンテンツ Tennis に対する注視領域 (紫領域) の変化 (被験者 H)。

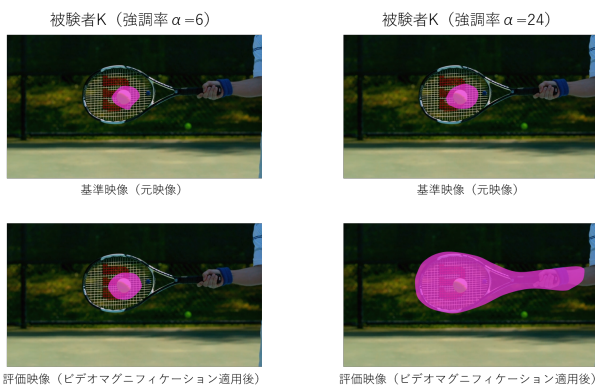


図 5: コンテンツ Tennis に対する注視領域 (紫領域) の変化 (被験者 K)。

強調率が大きくなるにしたがって、映像変化知覚が大きくなっていく傾向が見られ、線形回帰で分析を行った結果、映像変化知覚は強調率に比例する形で線形に有意に増加することが確認された ($p < 0.05$)。その一方で、強調率が大きくなるにしたがって、強調された映像に対する自然さは損なわれ、線形回帰で分析を行ってみると、線形に減少することが確認された ($p < 0.05$)。

5.2 ユーザーの注視領域の変化

図 4, 5 はコンテンツ Tennis に対してビデオマグニフィケーションを適用した際の被験者の注視領域 (紫) の変化を示した図である。この 2 つの図を見てみると、強調率が低い際は視覚的顕著性の高いと考えられる映像領域に注視領域が集中している。一方で、強調率が大きくなるにしたがって、ボール打撃時の衝撃がラケットや腕に伝播していく微細な物理現象が強調・可視化されるので、それに伴い注視領域がラケット全体や腕に広がっていく様子が確認された。この傾向は他コンテンツでも同様であった。

6. 考察

全体的な傾向として、ビデオマグニフィケーションの強調率を大きくすることで、ユーザーの映像変化知覚が増加

し、映像中に潜む微細な物理現象に気づきやすくなることから、ビデオマグニフィケーションという技術がこれまでヒトの視覚では捉えることが困難であった事象を伝えるのに有効であることが示唆された。

その一方で、微細変化の強調率の上昇にしたがって、基準映像に比べて評価映像の不自然さを強く感じるという結果が得られた。今回の主観評価実験では、基準映像と評価映像を対で視聴するような実験プロトコルを設定したため、基準映像にはない変化そのものを不自然であると感じ評価した可能性が考えられ、今後は「ビデオマグニフィケーションにとっての不自然さ」を定義する形容詞の選定が課題である。また、単一の映像で評価を行う ACR (Absolute Category Rating Method) 法で評価することで、対比較で生じた影響を多少軽減できるのではないかと考えている。

また、ビデオマグニフィケーションの強調率の上昇に伴いユーザーの注視領域が変化し、微細な物理現象が潜むと考えられる領域を注視するようになったことから、ビデオマグニフィケーションの本来の目的である「映像中に潜む微細な変化を可視化する」という技術の有用性を、映像変化知覚という量的側面と注視領域という空間的側面の両方から確認することができた。この結果から、ビデオマグニフィケーションを用いて注視領域を誘導する新しいユーザーインターフェースを研究開発できるのではないかと考えている。

7. まとめ

本稿では、ビデオマグニフィケーションという技術を用いて映像中の微細な変化を強調することで、ユーザーに与える映像変化知覚と注視領域の影響について評価・分析を行った。主観評価実験の結果、強調率を大きくすることでユーザーの映像変化知覚が増加し、注視領域が変化したことで、映像中に潜む微細な物理現象に気づきやすくなることが分かり、ビデオマグニフィケーションがこれまでヒトが捉えられていなかった微細な物理現象を伝えるのに有効であることが確認された。今後は、微細変化の強調によって生じる不自然さの評価手法の再検討および評価、主観評価に基づく強調率の最適化など課題として取り組んでいく。

参考文献

- [1] Wu, H.-Y., Rubinstein, M., Shih, E., Gutttag, J., Durand, F. and Freeman, W.: Eulerian Video Magnification for Revealing Subtle Changes in the World, *SIGGRAPH* (2012).
- [2] Wadhwa, N., Rubinstein, M., Durand, F. and Freeman, W. T.: Phase-based video motion processing, *SIGGRAPH* (2013).
- [3] Wadhwa, N., Rubinstein, M., Durand, F. and Freeman, W. T.: Riesz pyramids for fast phase-based video magnification, *The IEEE International Conference on Computational Photography (ICCP)* (2014).

- [4] Elgharib, M. A., Hefeeda, M., Durand, F. and Freeman, W. T.: Video magnification in presence of large motions, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2015).
- [5] Kooij, J. F. and van Gemert, J. C.: Depth-aware motion magnification, *The European Conference on Computer Vision (ECCV)* (2016).
- [6] Zhang, Y., Pinteá, S. L. and van Gemert, J. C.: Video Acceleration Magnification, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).
- [7] Takeda, S., Okami, K., Mikami, D., Isogai, M. and Kimata, H.: Jerk-Aware Video Acceleration Magnification, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).
- [8] Fleet, D. J. and Jepson, A. D.: Computation of component image velocity from local phase information, *International journal of computer vision*, Vol. 5, No. 1, pp. 77–104 (1990).
- [9] Hu, Y., Song, R. and Li, Y.: Efficient coarse-to-fine patchmatch for large displacement optical flow (2016).
- [10] ITU-T Recommendation P.910: Subjective video quality assessment methods for multimedia applications (1999).
- [11] ITU-R Recommendation BT-500-11: Methodology for the subjective assessment of the quality of television pictures (2002).