

平成 30 年 6 月 12 日現在

機関番号：13401

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K00150

研究課題名(和文) 4K UHD画像符号化における時空間の最適レート制御に関する研究

研究課題名(英文) Research on optimal spatio-temporal rate control for 4K UHD videos

研究代表者

吉田 俊之 (YOSHIDA, Toshiyuki)

福井大学・学術研究院工学系部門・教授

研究者番号：50240297

交付決定額(研究期間全体)：(直接経費) 2,500,000円

研究成果の概要(和文)：4K60P動画の伝送には、効率的な圧縮符号化法と共に符号化画質を最大化する時空間のレート制御手法が不可欠である。本研究では、代表者が過去に確立した「主観評価値の予測値(EMOS)を最大化する時空間レート制御手法」の4K60P動画への拡張を目指した。まず、4K60P動画に対する時空間EMOSの予測手法を確立した後に、これをH.265/HEVC符号化器に実装し、EMOS最大化符号化を実現した。さらに、定EMOS符号化や画質平坦化手法への応用を通して提案手法の有効性を確認すると共に、60Hzのフレームレートでは4K解像度を十分に活用できない場合があることを実験的に明らかにした。

研究成果の概要(英文)：4K60P video transmission requires an efficient video encoder as well as a spatio-temporal rate control technique that maximizes resulting video qualities. This research has aimed at an extension of such a rate control technique, previously proposed by the author, into 4K60P videos, i.e., the rate control technique based on a maximization of an estimated subjective video quality (EMOS). A spatio-temporal EMOS estimation technique for 4K60P videos was first established, and then it was applied to H.265/HEVC encoder to realize an EMOS-maximization-based video encoder. It was also applied to a constant-EMOS coding and a video quality equalization technique to verify the validity of our approach. These experiments have suggested that a frame rate of 60Hz is often insufficient to make the best of the 4K video resolution.

研究分野：画像処理・画像符号化

キーワード：動画符号化 4K動画 時空間画質 MOS 画質制御 レート制御 定MOS符号化

## 1. 研究開始当初の背景

4K 動画は、画素数  $3860 \times 2160$ 、フレームレート  $60 \text{ frame/s}$  (fps) を代表例とする次世代の映像フォーマットである。その情報は非圧縮状態で  $10 \text{ Gbit/sec}$  (bps) にも昇るため、その伝送や蓄積には高効率の圧縮符号化が必要となる。

4K 動画の圧縮符号化は、これまで同様、原画像-復号画像間の平均 2 乗誤差 (MSE) を最小化する固定帯域符号化を指向して研究が進められてきた。一方で、4K 動画は今後の急速な普及が予想され、現状のインターネットのような伝送帯域が刻々変動するネットワーク上での伝送や、モバイル環境下で小規模のメモリ媒体に保存する等の応用が進むと考えられる。その際は、従来の固定帯域を指向した圧縮符号化よりも、更なる高画質性と柔軟性を備えた圧縮符号化が求められる。

一般に、圧縮符号化における画質の評価は客観評価と主観評価に大別され、前者は MSE に代表される符号化前後の数値指標に基づく手法、後者は人間が知覚する画質を数量化した主観評価値に基づく手法である。画質評価は本来的には主観評価に基づくべきであるが、客観評価は符号化アルゴリズムとの親和性が高く、設計や解析が容易となるため現状では MSE に基づく客観評価が広く用いられている。しかしながら、異なる画像間では MSE-主観画質間の相関は非常に低く、MSE 最小規範の下での符号化画質は必ずしも最大化されないことは周知の事実である。また、符号化動画の画質は「動きの滑らかさ」と「各単一フレームの画質」によって定まるため、従来のようにフレームレート一定の下で各フレームの画質を最大化しても時空間の総合画質は最大化されず、柔軟性の高い符号化法とは言えない。

この問題に対して研究代表者は、代表的な主観画質評価値である MOS (Mean Opinion Score) を動画シーン単位として推定し、これを最大化する手法を提案してきた [1-4]。この手法では、動画の符号化ビットレート (BR) を  $R$  [bps]、毎秒のフレームレート (FPS) を  $t$  [fps]、フレーム当たりの平均ビット量 (BPF) を  $s$  [bits/frame (bpf)] としたとき、

$$R = s \cdot t \quad (1)$$

と書くことに着目し、図 1 のように対象動画を  $(s, t)$  の下で符号化した場合の MOS (5 段

階スケール) の推定値  $EMOS(s, t)$  を求め、これを最大化する  $(s, t)$  の下で符号化を行なう。

文献 [1] では、 $EMOS(s, t)$  を

$$EMOS(s, t) \propto EMOS_s(s) \cdot EMOS_t(t) \quad (2)$$

のように、与えられた  $s$  に対して単一フレーム内の空間画質を表す  $EMOS_s(s)$ 、および与えられた  $t$  に対して時間方向の動きの滑らかさを表す  $EMOS_t(t)$  に分離して推定可能なことを確認し、文献 [2] 等では  $EMOS_s(s)$  と  $EMOS_t(t)$  を対象動画の特徴量から推定する手法を与えている。対象各シーンを単位として  $EMOS(s, t)$  が推定できれば、図 1 に示すように、与えられた BR  $R$  によって定まる双曲線上で  $EMOS(s, t)$  が最大となる  $t$  (最適 FPS) および対応する  $s$  の組合せの下で符号化すれば、限られた BR 制約の下で画質を最大化する符号化が実現できる。文献 [3] では、このような時空間レート制御法の実装の詳細を論じている。

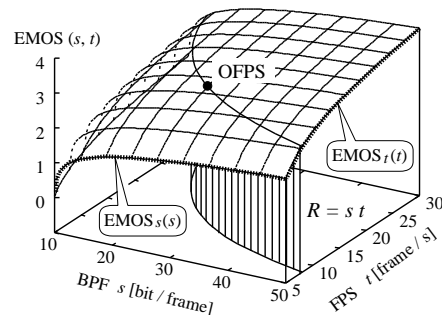


図 1 :  $EMOS(s, t)$  の例と最適 FPS の決定

以上の時空間レート制御法は、これまで SIF ( $352 \times 240$ ) サイズから 2K ( $1920 \times 1080$ ) サイズへと拡張されてきたが、4K60P 動画への適用については未だ検討されていない。そこで本研究では、4K60P 動画を対象として、従来の MSE に基づく画質/レート制御ではなく、より柔軟な EMOS ベースの制御手法の確立、応用、検証を目指して研究を行なった。

## 2. 研究の目的

以上の背景に基づき、本研究では、以下の 3 項目を研究の目的とした。

- (1) 1. で述べた時空間レート制御法を 4K60P 動画に拡張するため、式 (2) に基づく時空間の EMOS 予測手法を確立する
- (2) (1) の EMOS 予測手法に基づく時空間レート制御法を実現し、H.265/HEVC 符号化器上に実装する

- (3) 応用として、最適 FPS 下の符号化、定 EMOS 符号化、画質平坦化法 [4] に適用し、提案手法の妥当性を検証する

### 3. 研究の方法

研究の目的 (1)–(3) に対応し、以下の方法に従って研究を進めた。

#### (1) 4K60P 動画像に対する EMOS 予測

##### ① 空間方向の $EMOS_s(s)$ の予測

研究代表者らが文献 [2] で提案した SIF サイズの動画像に対する EMOS 予測法を 4K60P に拡張する。文献 [2] では、原画像–劣化画像間の PSNR [dB] 値と MOS の間には強い相関があることを利用し、空間 EMOS 予測を「(a)  $s$  から PSNR の予測」と「(b) PSNR から EMOS の予測」に分離して  $EMOS_s(s)$  を予測している。本研究においても  $EMOS_s(s)$  の予測を同様に分離し、(b) については 4K60P 試験画像に対する主観評価実験に基づいて予測式を導出する。一方の (a) については、従来手法よりも計算負荷が小さい手法を検討する。

##### ② 時間方向の $EMOS_t(t)$ の予測

動画像における時間方向の動きの滑らかさは、フレームレートおよび動領域と静止領域の相対関係に基づいて知覚されると考えられる。そこで本研究では、動領域と静止領域を抽象化した人工動画像を生成し、その実測 MOS のパラメトリック近似に基づいて  $EMOS_t(t)$  の予測式を導出する。具体的には動領域のサイズと移動速度、フレームレートが相異なる複数の人工動画像を生成して主観評価実験により MOS を測定し、動画像から抽出される特徴量を用いてパラメトリック近似することにより予測式を導く。

#### (2) H.265/HEVC 符号化器上への実装

(1) で得られた EMOS 予測手法およびそれに基づくシーン単位の時空間レート制御法を H.265/HEVC 参照ソフトウェア上に実装する。

#### (3) 応用と検証

(2) で実装した時空間レート制御法/符号化器を、最適 FPS 下での CBR(Constant Bit Rate) 符号化、一定 EMOS を維持する VBR(Variable —) 符号化(定 EMOS 符号化)、および大規模バッファを利用して各シーンの画質

の平坦化を図る疑似 CBR 符号化へ応用すると共に、各手法に関する有効性の検証を行なう。

### 4. 研究成果

本研究では、複数の 4K60P 試験画像を利用して 3. の項目 (1)–(3) を実施した。

#### (1) 4K60P 動画像に対する EMOS 予測

##### ① 空間方向の $EMOS_s(s)$ の予測

単一シーンから成る 15 種類の 4K60P 試験画像に対して適当な BPF  $s$  を設定して HEVC 符号化し、主観評価実験によって MOS (1~5 の 5 段階評価) を測定すると共に PSNR のシーン内平均値を求め、文献 [2] の手法を用いて PSNR から EMOS を予測する予測式

$$\begin{aligned} EMOS(x) &= aT(x-T)^3 + bT(x-T) + c \quad (3) \\ T &= c_0 + c_1 \ln(\varepsilon + 1) + \\ &\quad c_2 \ln(E + 1) + c_3 \ln(L + 1) \quad (4) \\ a &= 1.122 \times 10^{-5} \\ b &= 9.323 \times 10^{-4} \\ c &= 5.000 \\ c_0 &= 85.60 \\ c_1 &= -4.384 \\ c_2 &= -1.785 \\ c_3 &= -2.892 \end{aligned}$$

を導出した。ここで、 $x$  は対象画像–原画像間の PSNR [dB] のシーン内平均、 $\varepsilon$ 、 $E$ 、 $L$  は対象画像から抽出される特徴量で、各々、

$\varepsilon$  : Sobel オペレータ出力のフレーム間平均  
 $E$  : フレーム間動き推定残差のフレーム間平均  
 $L$  : フレーム内平均輝度値のフレーム間平均

である [主な発表論文等の雑誌論文 ①]。15 種類の試験画像から順次 1 種類をオープンデータとする leave-one-out 法により式 (3) の推定精度を検証したところ、平均誤差 0.38、最大誤差 1.08 が得られ、ほぼ実用的な精度で PSNR から MOS が推定可能なことを確認した。

一方、BPF  $s$  から PSNR  $x$  を予測する手法として、文献 [2] では相異なる 4 点の  $s$  (RD ポイント) で符号化した PSNR を曲線近似する手法を用いていたが、4K 画像に対しては符号化の計算負荷を抑えるため、より少数の RD ポイントで予測可能な手法を検討した。まず、 $s$ -PSNR  $x$  の関係を近似するモデルとして、

$$x = a(s - c)^b \quad (5)$$

を用いる。与えられた動画像の RD ポイントから定数  $a$ 、 $b$ 、 $c$  が推定できれば、 $s$  から PSNR

$x$ が予測できる．そこで，18種類の4K60P試験画像を各5点のRDポイントでHEVC符号化した $s$ - $x$ の関係を，各試験画像毎に式(5)で近似して定数 $a$ ， $b$ ， $c$ を求めたところ，各 $a$ ， $b$ ， $c$ は $s = 0.083$  [Mbps]でHEVC符号化した際のPSNR  $p$  [dB]を用いて高精度に近似表現可能なことが判った．実際，

$$\begin{aligned} a &= a_0 \cdot (p - a_1)^{a_2} \\ b &= b_0 \cdot (p - b_1)^{b_2} + b_3 \\ c &= c_2 \cdot \tanh(c_0 \cdot (p - c_1)) - c_3 \end{aligned} \quad (6)$$

で近似し，最小二乗法によって

$$\begin{aligned} a_0 &= 30.13, a_1 = 24.06, a_2 = 0.1678 \\ b_0 &= -0.003406, b_1 = 28.45 \\ b_2 &= 1.109, b_3 = 0.1174 \\ c_0 &= 3.585, c_1 = 34.04 \\ c_2 &= 0.003829, c_3 = -0.01630 \end{aligned}$$

を得た．与えられた動画像に対しては， $s = 0.083$  [Mbps]でHEVC符号化したPSNR  $p$ を式(6)に代入して $a \sim c$ を求め，これを式(5)に代入することで任意の $s$ に対するPSNR  $x$ を推定できる．最終的に式(3)と併用することにより， $EMOS_s(s)$ が得られる．

## ② 時間方向の $EMOS_t(t)$ の予測

本研究では，液晶モニタ上の倍速表示等の付加機能はすべて解除して用いる．まず，フレームレートを変化させた複数の4K60P試験画像を複数の液晶モニタ上に表示し，黒挿入等の表示方式の違いによるMOSへの影響を事前調査した．その結果，表示方式の違いが時間方向のMOS，すなわち動きの滑らかさに与える影響は，想定されるEMOSの推定誤差に比べて小さく，無視し得ると結論した．

$EMOS_t(t)$ の予測式の導出には，画像内の動領域と静止領域をモデル化した人工動画像を用い，動領域の面積と移動速度およびフレームレート(60, 30, 15fps)を変化させた計60種類を生成してMOS(1~5の5段階評価)を測定した．当初は人工画像のみで予測式の導出を試みたが，後述する特徴量の分布に偏りが生じたため，フレームレート60, 30, 15fpsの計36種類の自然動画像に対して測定したMOSも併用した．

ここでは特徴量として，

$d$  : フレーム間差分のシーン平均

$a$  : 動領域の画素数のシーン平均 [画素]

を用いて予測を行なう．前者の $d$ は，隣接フレーム輝度の画素単位の差分に対して $T = 13$ 以上で閾値処理し，フレーム内総画素数に対して $2/10000$ 以下の小領域を除去した後，シーン全体で平均を取ることによって求め，後者の $a$ は， $d$ を求める際に有意な差分を有する総画素のシーン平均とする．各フレームレート毎の人工および自然動画像から特徴量 $d$ と $a$ を抽出した後，EMOSと $t$ ， $d$ ， $a$ の関係をモデル

$$EMOS_t(t) = \frac{1}{e^{f(t,d,a)} + 1} + 1 \quad (7)$$

を用いて近似する．ここで， $f(t, d, a)$ は，

$$f(t, d, a) = A \cdot \log d + B \cdot \log a + \frac{C}{t^2} + D \quad (8)$$

によって予測する．定数 $A \sim D$ を重回帰分析によって求めたところ，

$$\begin{aligned} A &= 0.5789, B = -1.5005 \\ C &= 619.9232, D = 1.3383 \end{aligned}$$

が得られた．これらの定数，対象動画像から抽出した $d$ と $a$ ，およびフレームレート $t$ を式(7)，(8)に代入することにより，時間方向のEMOSが推定される．

式(7)，(8)の妥当性を自然画像に対するleave-one-out法により検証したところ，平均誤差0.29，最大誤差1.05が得られた．現段階では，最大誤差が若干大きい一方で平均誤差は十分小さく，実用上十分な精度を有する予測式が導出されたと判断できる．

## (2) H.265/HEVC符号化器上への実装

(1)によって4K60P動画像に対する時空間のEMOS予測が可能となったため，これに基づく時空間のレート制御法をHEVC符号化器上に実装した．これはマルチパスの符号化器で，初段のパスで対象動画像からMOS予測に必要なシーン単位の特徴量を抽出して $EMOS(s, t)$ を予測し，これを元に時空間レート制御を実行した後，次段のパスで実際の符号化を行なう．

## (3) 応用と検証

### ① 最適FPS下でのCBR符号化

最初の応用として，4シーン(SC)から成る4K60P動画像360フレームに対して一定BRを与えて各SCの最適FPSを予測し，その下でCBR符号化を行なった．図2は各SCに対して予測された $EMOS(s, t)$ の分布，表1は定

ビットレート  $R = 15$  [Mbps] を与えた場合の各 SC の最適 FPS および対応する EMOS である。

表 1 において、最大 FPS を対象動画と同じ 60fps とした場合、最適 FPS はすべて 60fps が選択される一方、SC1 と 4 の EMOS は低い値に留まっている。これは図 2(b), (d) において、時間方向の EMOS が 60fps で飽和に達していないため、4K 画像によっては 60fps では不十分で、より高い FPS が必要となる場合があることを示唆している。そこで、式 (7), (8) の予測式が 60fps を越える 120fps まで適用可能であると仮定して最大 FPS を 120fps まで許容した場合には、表 1 に示すように SC2 と SC4 では最適 FPS として 120fps が選択され、対応する EMOS も改善されている。

以上より、提案手法によって SC 毎の最適 FPS が推定可能であること、また 4K の解像度をフルに活用するためには 60fps を越える 120fps 程度の FPS が必要となる場合があることが結論される。

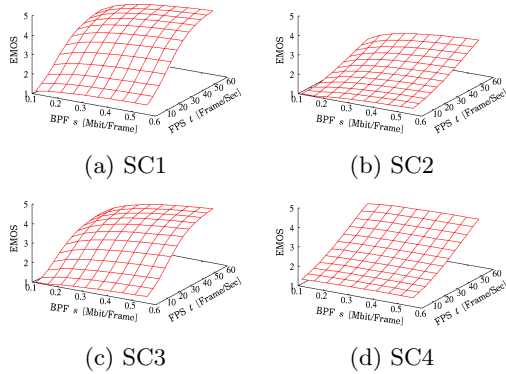


図 2 : 各 SC に対する  $EMOS(s, t)$

表 1 : 各 SC に対する最適 FPS と対応する EMOS

SC	最大 FPS 60fps		最大 FPS 120fps	
	最適 FPS	EMOS	最適 FPS	EMOS
1	60fps	3.89	60fps	3.89
2	60fps	3.13	120fps	3.54
3	60fps	3.75	60fps	3.75
4	60fps	3.06	120fps	4.20

## ② 定 EMOS 符号化

(2) で実現した符号化器を用いると、副産物として SC 単位の EMOS が得られるため、各 EMOS が一定となるようにレート制御して VBR 符号化することで定 EMOS 符号化が実現できる。ここでは、① の 4 つの SC に対し

て 120fps を許容し、EMOS = 3 および 4 を与える定 EMOS 符号化を行なった。表 2 は、その際に必要となる各 SC の BR である。

最適 FPS 下の CBR 符号化では SC 間での EMOS の変動が大きい一方、定 EMOS 符号化では EMOS は平坦化される反面、各 SC で要求される BR の変動は数倍程度と非常に大きくなる。

表 2 : 各 SC に対して指定 EMOS を与える BR

SC	EMOS = 3	EMOS = 4
1	6.0Mbps	18.0Mbps
2	9.6Mbps	26.4Mbps
3	4.8Mbps	27.0Mbps
4	3.6Mbps	9.6Mbps

## ③ 画質平坦化

一般に、複数 SC から成る動画の画質は、MOS 最小 SC の画質に支配される傾向があるため、与えられた BR  $R$  の下で実現される動画全体の画質を最大化するためには、最小 MOS の改善が必要となる。そこで文献 [4] では、大容量のバッファを用い、与えられた一定の平均伝送 BR  $R$  の下で符号化レートを SC 間で融通し合うことで EMOS の平坦化を図る手法 (画質平坦化法) が提案されている。

本手法では、まず利用するバッファ量  $B_{max}$  (単位: 秒) とその初期占有量、画質平坦化の過程で各 SC が選択可能な BR (離散値) を規定する。いま、対象区間に  $i = 1, 2, \dots, n$  の  $n$  個の SC があると仮定し、 $r_i$  を各 SC に与える BR として、

$$\max_{r_1, r_2, \dots, r_n} \min_{1 \leq i \leq n} EMOS_i(r_i) \quad (9)$$

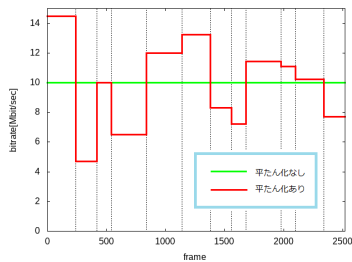
$$\text{s.t.} \quad \sum_{i=1}^n r_i T_i \leq R \sum_{i=1}^n T_i$$

$$0 \leq B_i \leq B_{max} \quad (1 \leq i \leq N)$$

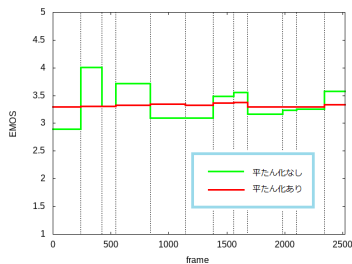
に基づく BR 割り当てによって平坦化を行う。ここで、 $T_i$  は SC  $i$  の持続時間、 $B_i$  は SC  $i$  の符号化終了時点でのバッファ占有量 (単位: 秒) である。実際には、初期条件として各 SC に対して平均伝送 BR  $R$  を与えた後、最も高い EMOS を与える SC の BR を 1 段階下げ、その過程で生じた余剰ビットを最小 EMOS を有するシーンに割り当てる処理を繰り返すことで、式 (9) を近似的に実行する。

本手法を、長さ 2520 フレームで 42 の SC から成る 4K60P 動画に適用し、画質平坦化の有

効性を検証した。平均伝送BRを10Mbps、各SCが選択可能なBRを0.6~36 [Mbps]、バッファ量  $B_{max}$  を42秒 (420Mbit) に設定し、本手法を適用した結果を図3に示す。図3(a)はSCを単位とするBRの推移、(b)はEMOSの変動を表しており、緑線が平坦化前、赤線が平坦化後を表す。本手法によってSC間でBRが融通され、最小EMOSが改善されて変動幅が低減されていることが確認できる。



(a) 平坦化前後のBRの変動



(b) 平坦化前後のEMOSの変動

図3：画質平坦化法の適用結果

以上、本研究はほぼ当初目標の通り実施され、所期の成果が得られたと結論される。

#### <引用文献>

- [1] 稲積泰宏 他, “ビットレート制限下における動画通信のための最適フレームレート推定”, 電子情報通信学会論文誌 (B), vol.85-B, no.7, pp.1130-1142 (2002年7月)
- [2] 宮田高道 他, “SNRに基づく動画の主観画質推定”, 電子情報通信学会論文誌 (A), vol.J88-A, no.11, pp.1292-1296 (2005年11月)
- [3] 吉田俊之 他, “主観評価値の最大化に基づく動画の時空間レート制御とその符号化への応用”, 映像情報メディア学会論文誌, vol.62, no.11, pp.1785-1793 (2008年11月)
- [4] 吉田俊之, “予測MOSの平坦化に基づく符号化画質改善法”, 電子情報通信学会

論文誌 (D), vol.J93-D, no.9, pp.1634-1637 (2010年9月)

## 5. 主な発表論文等

[雑誌論文] (計1件)

- ① 吉田 俊之, “4K圧縮動画に対する主観画質推定”, 電子情報通信学会論文誌 D, 査読有, vol.J100-D, no.9, pp.835-838, 2017年

[学会発表] (計5件)

- ① Akira Yamanaka and Toshiyuki Yoshida, “EMOS-based spatio-temporal rate control technique for 4K60P videos”, International Workshop on Advanced Image Technology 2018, Poster Session 1-22, Jan. 7-9, 2018, Chiang Mai(Thailand)
- ② 山中 彰, 吉田 俊之, “4K60P動画に対するEMOS最大化レート制御について”, 第32回画像符号化シンポジウム, P-1-15, 2017年11月20-22日, ラフォーレ修善寺 (静岡県伊豆の国市)
- ③ 山口 翔也, 吉田俊之, “4K動画を対象とした時間方向のMOS予測手法について”, 平成29年度電気関係学会北陸支部連合大会, F2-35, 2017年9月11日, 富山大学 (富山県富山市)
- ④ 山中彰, 吉田俊之, “1点のRDポイントの計測に基づくRD曲線の高精度予測について”, 第31回画像符号化シンポジウム, P-4-07, 2016年11月16-18日, ラフォーレ修善寺 (静岡県伊豆の国市)
- ⑤ 菊野哲誠, 吉田俊之, “4K60P動画に対する時空間のMOS予測について”, 第31回画像符号化シンポジウム, P-4-09, 2016年11月16-18日, ラフォーレ修善寺 (静岡県伊豆の国市)

## 6. 研究組織

- (1) 研究代表者  
吉田 俊之 (YOSHIDA Toshiyuki)  
福井大学・学術研究院工学系部門・教授  
研究者番号：50240297
- (2) 研究分担者  
なし
- (3) 連絡研究者  
なし