

K-001

## ハンドサインを記録した一人称視点映像からのノウハウ映像の生成 Generation of know-how video from first person perspective video recording hand signs

伊関 信之<sup>†</sup> 嶋田 聡<sup>†</sup>  
Nobuyuki Iseki Satoshi Shimada

### 1. はじめに

技能伝承やスキル獲得等の学習には映像がよく利用されている[1]。有効な教材映像を制作するには適切な素材映像の撮影および編集が必要なので自前で制作するには負荷が大きい。撮影の負荷については一人称視点映像を用いることで解決できる。著者らは一人称視点映像を有効に活用できる方法として手の平の重ね書きで映像シーンに対してメモを記録する映像へのアノテーション方法を検討している[2]。この方法はメモの記録に時間がかかるが、映像編集に特化したハンドサインであれば作業に実施や見学中に付与することができる。

本研究では、作業に立ち会う人の頭部に装着したカメラで撮影する一人称視点映像で素材映像を取得し、編集映像に採用する有効区間を示すハンドサインをその場で記録しておくことでノウハウ映像を効率よく編集する方法を検討する。

### 2. 提案方法

#### 2.1 一人称視点映像の編集

教材映像の制作では死角がなく作業内容を把握しやすいアングルで撮影した素材映像の取得が重要である。本方法では、作業内容が常時記録されている作業者の視点映像と、見やすいアングルに自然と移動して作業に注視する傾向がある見学者の視点映像を素材映像とする。撮影した素材映像を後から編集するのは大変なので、作業中に作業者と見学者に、編集映像に採用する有効区間の in と out をハンドサインで自身の視点映像に記録してもらう。

映像編集については、視点が安定している見学者の視点映像を優先的に利用することとする。また、in, out のタイミングを修正(変更)したいこともある。in のサインは未来の予測で出すので、修正の頻度が高いと考えられる。修正を簡易に行えるように連続して in を出すことを許容し、最後に出した in を有効とする。out については作業結果を把握したうえで出すので修正はなしとし、誤って out のサイ

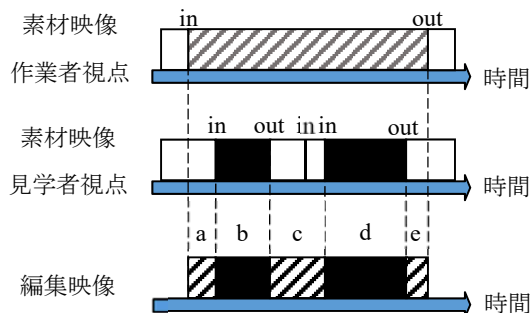


図1 編集方法の考え方

<sup>†</sup> 日本大学工学部

ンを出した場合には直後に改めて in, out の有効区間を指定することとする。従って、図1に示すようにハンドサインを出した場合には、映像編集の区間 a と e は作業者のみ有効区間なので作業者視点映像が、区間 b と d は見学者を優先して見学者視点映像が、区間 c は連続して出された in の最後を有効とした見学者視点映像が、それぞれ採用される。

#### 2.2 ハンドサイン

有効区間の in と out を指定するハンドサインは出しやすく、記憶しやすいものが求められる。また、ハンドサインを含む映像からサインを出している区間の検出のハンドサインの認識が容易に行えるようにしたい。この両方を満足するハンドサインの方法として、手の形状で in と out を区別することとし、その前後で共通の手の動きを行うこととする。ハンドサインの出し方は、図2に示すように、①手を胸から視界の中央に向けて上げ、次に、②手を注視して2秒間静止させ、最後に、③手を下げるものとする。手の形状は in のときはパー、out のときはグーとする。

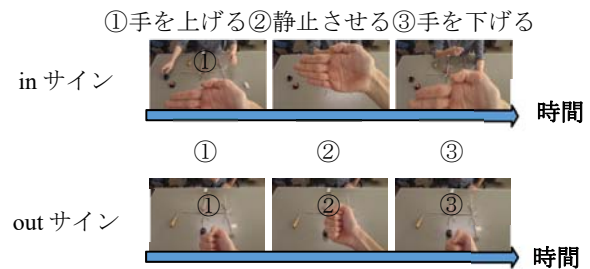


図2 ハンドサインの動作

### 3. 編集区間の自動検出

前述の提案方法でハンドサインの記録を含む映像を取得した時に視点映像から編集区間情報を自動検出する方法を図3に示す。まず Step1 では手の動きからハンドサインの候補区間を検出する。次に Step2 では画像内容を認識して「in のサイン」、「out のサイン」、「どちらでもない」のいずれかに判別する。最後に Step3 で in と out の区間から

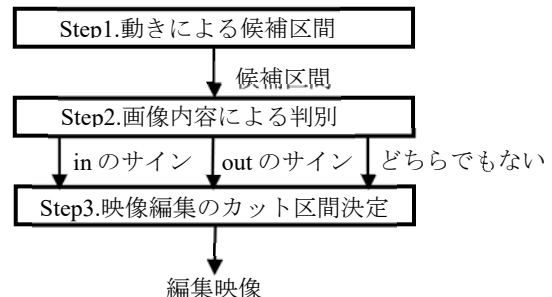


図3 提案方法における編集区間情報の生成手順

事前に構築したルールに基づいて、編集映像に取り込む有効区間を決定する。

### 3.1 Step1 の動きによる候補区間の検出方法

ハンドサインに相当する手の動きは、画像を  $N \times N$  のブロックに分割し、手の写っている中央付近のブロックの動きベクトルから検出する。ハンドサインには①手を上げる、②静止させる、③手を下げる、の 3 つ順に動きのパターンがある。動きベクトルの大きさが閾値以上の画素について、それらの画素がブロックに占める割合（動き量と呼ぶ）、動きベクトルの方向の平均と標準偏差の 3 つの特徴量からハンドサインの動きが発生したと考えられる区間を候補区間とする。

### 3.2 Step2 の画像内容による判別方法

検出した候補区間において in または out の手形状の物体があるかを R-CNN で判別する。ハンドサインの候補区間として検出された区間にある静止区間の開始から終了までの間にハンドサイン判別の信頼度を表すスコアがしきい値以上のフレーム画像が P フレーム以上連続し、ハンドサインの in と out の種類の藩閥結果が一致したものを正検出とする。

## 4. 実験

### 4.1 方法

約 5 分の組み立て作業を、作業者の視点映像と見学者の視点映像を各 1 名で 5 回、撮影した。撮影映像の例を図 4 に示す。被験者は 4 名である。R-CNN の識別器の学習は、被験者 4 人の中の 1 人に対して in と out のハンドサインを行っているときの画像を各 200 件、別に取得して行った。



図 4 取得した素材映像

### 4.2 結果

Step1 の候補区間の検出で抽出した 3 つの特徴量を図 5 に示す。ハンドサインの動きに対応して変化していることがわかる。Step2 の判別については図 6 に示すように「in のサイン」、「out のサイン」の判別が行われた。「in のサイン」、「out のサイン」の画像に対する検出信頼度を表すスコアは 0.9 と高くなっている。このスコアの閾値処理で「どちらでもない」を判定できる。

ハンドサインの検出結果を表 1 に示す。適合率は正検出数/検出数、再現率は正検出数/サイン数である。被験者 4 名の in と out のハンドサインはそれぞれ、in のときは 53 件、out のときは 48 件で、適合率は in が 95.7%、out が 97.7% と高く、再現率は in が 83.0%、out が 87.5% であった。高い適合率から提案方法の有効性が確認された。再現率が低くなったのは step1 の候補区間検出が原因で、手の静止区間が

未検出になったためである。今後は、視点変動によるグローバルモーションを考慮して手の動き量を評価する方法や、より安定して検出できるハンドサインの提示方法について検討する予定である。

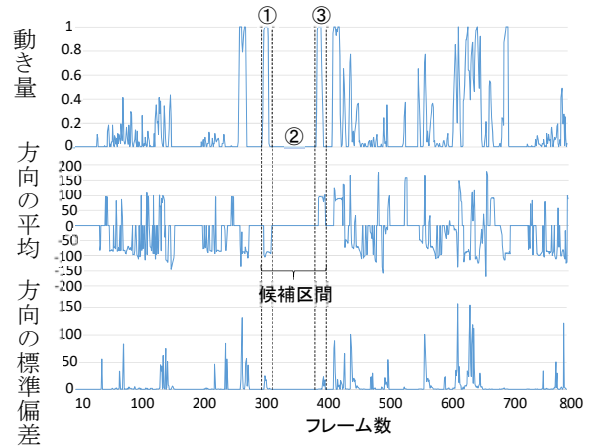


図 5 動きベクトルによる候補区間の検出例



in のサイン  
スコア : 0.93

out のサイン  
スコア : 0.95

図 6 画像内容による判別結果の例

表 1 ハンドサインの検出結果

	サイン数	検出数	正検出数	適合率	再現率
in	53	46	43	95.7%	83.0%
out	48	43	42	97.7%	87.5%

## 5. おわりに

一人称視点映像を素材映像とし、in と out のハンドサインによる映像編集方法について検討した。被験者 4 名の小規模な実験により、視点映像に記録したハンドサインの自動検出および判別を行えることを確認した。今後は実験映像を増やし、提案方法の有効性を検証する。

### 参考文献

- [1] 畠田, “映像コミュニケーションによるスポーツスキルの獲得支援”, 映像学会, Vol.70, No.5, pp.725-728(2016).
- [2] 畠田,小南, “カメラを覆う動作と手の平の重ね書きによる一人称視点映像へのアノテーション”, 電気学会論文誌, Vol.137, No.9, pp.1238-1247 (2017).