

面接志願者の身体から生じた時系列変化を用いた 第一印象認識の設計

中井脩介 井上路子 米田駿介 西山正志（鳥取大学）



目的

面接志願者の全身動画から第一印象を精度よく認識するため
志願者の頭と上半身と下半身の身体部位の短動画をサンプリングし
時系列回帰を適用する手法を提案



背景:

- 採用面接で志願者は面接官へ良い第一印象を与える動作を行う
- 動作に関する第一印象の自動認識が求められている

仮説:

対面で行う面接では
志願者の全身が面接官の視界に入るため
全身の時系列変化が第一印象形成に重要

既存手法: **頭のみ**の時系列変化から特徴を抽出し印象を認識

[Kim+, IEEEAccess'23]
[Putra+, IEEEAccess'24]
[Le+, ICMI'24]

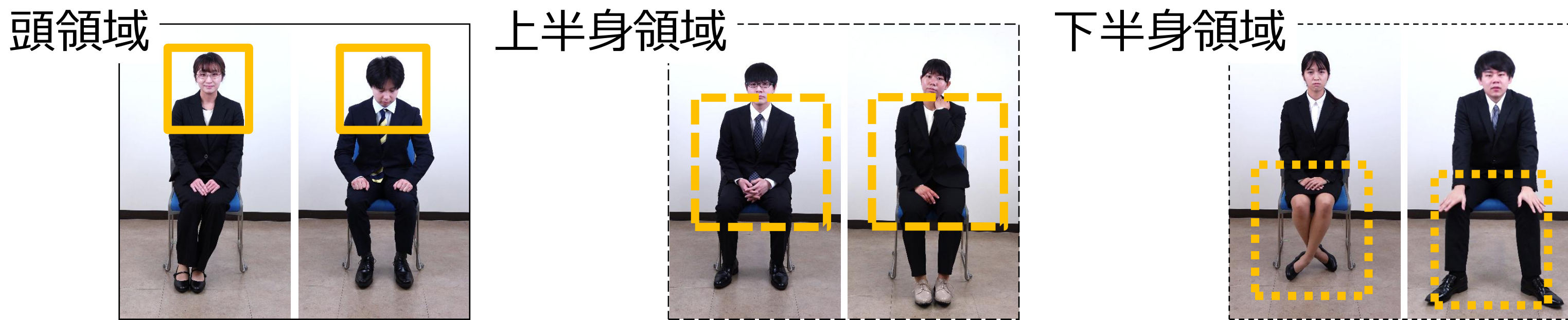
面接時に志願者が行う動作



時間方向に複雑に組み合わさった動作から第一印象が形成される

身体部位を含む各領域

多くの動作が出現する領域から特徴を抽出することを狙う



提案手法

- 特徴抽出のために頭領域・上半身領域・下半身領域の短動画に動画エンコーダを適用
- 時系列回帰のためにTransformerを適用

出力: 主観スコア(良い/悪い)

S4: 特徴間の時間方向の関係から時系列回帰

Transformerエンコーダ

空間方向に結合された各時刻の特徴を適用

S3: 同時刻の各短動画から抽出された特徴を頭と上半身と下半身との間で結合

S2: サンプリングされた短動画から特徴抽出

個々の短動画に動画エンコーダ(MViTv2)を適用 [Li+, CVPR'22]

動画エンコーダ

動画エンコーダ

動画エンコーダ

...

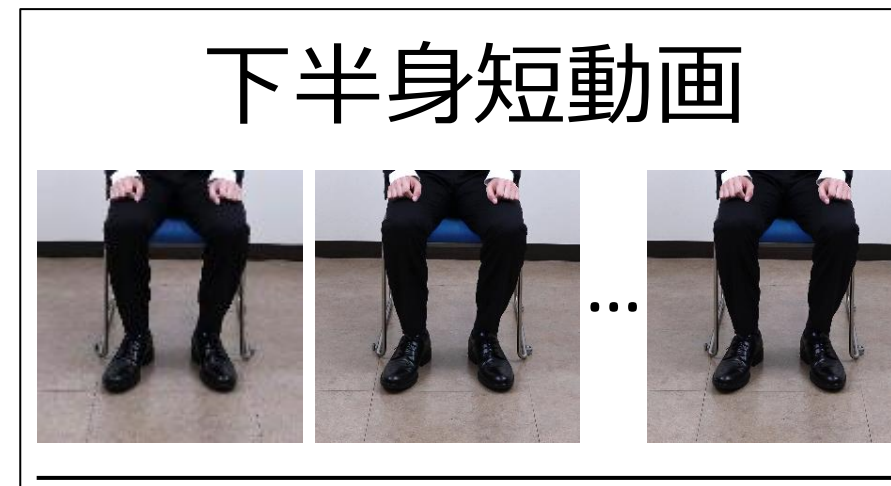
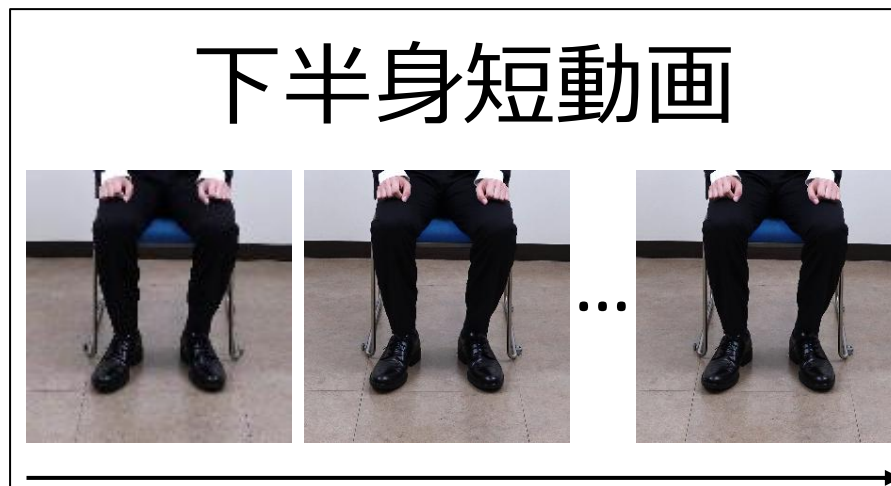
動画エンコーダ

動画エンコーダ

動画エンコーダ

S1: 頭と上半身と下半身の短動画を時空間グリッドサンプリング

短動画の時間長を1.8秒に設定



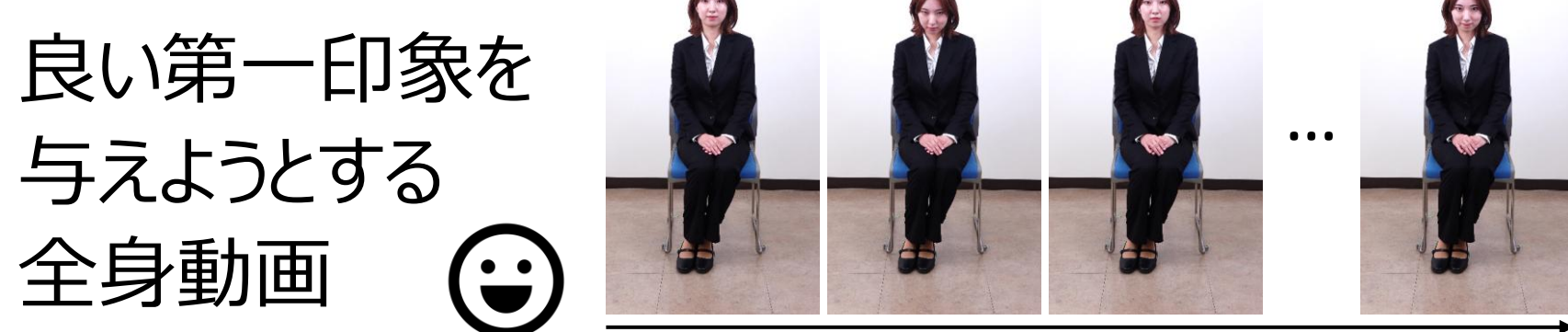
入力: 志願者の全身動画



実験

実験条件

志願者: 18名(男9名・女9名・平均年齢22歳)



動画数: 18(名)×2(良し悪し)=36(個)

印象評価者:

20名(男15名・女5名・平均年齢23歳)

主観スコアの付与:

良い/悪い第一印象を感じたかの主観スコアを
5段階で付与し全身動画ごとに平均を算出

評価方法:

- Leave-one-participant-out
- 予測された良い/悪い第一印象の主観スコアの
MAE(平均絶対誤差)を評価

実験1: 回帰で予測された主観スコアの誤差

手法	身体部位	特徴抽出	回帰手法	MAE↓
既存手法1 [Kim+, IEEEAccess'23]	頭	映像フレームの空間特徴マップ	LSTM	1.58
既存手法2 [Putra+, IEEEAccess'24]	頭	ランドマークとアクションユニットと視線と頭姿勢	Transformer	0.87
既存手法3 [Le+, ICMI'24]	頭	アクションユニットの時間特徴マップ	Transformer	1.58
提案手法	頭と上半身と下半身	短動画の時空間特徴マップ	Transformer	0.42

✓ 頭の特徴に加え上半身と下半身の特徴を第一印象認識に用いる有効性を確認

実験2: 提案手法のパラメータを変えた場合の誤差

身体部位	回帰手法				
	Averaging (短動画特徴の平均)	Linear (線形層1つ)	LSTM [Sak+, CoRR'14]	GRU [Chung+, CoRR'14]	Transformer [Vaswani+, CoRR'17]
頭	1.34	1.27	1.57	1.55	0.60
上半身	1.30	1.23	1.57	1.52	0.49
下半身	1.44	1.44	1.58	1.54	0.56
全身	1.40	1.37	1.58	1.56	0.45
頭と上半身と下半身	0.88	0.78	1.56	1.51	0.42

✓ 複数の身体部位を同時に用いつつTransformerを適用した方がよい

(MAE↓)