

鈴木研究室の紹介

データマイニング, 機械学習, AI

鈴木英之進

システム情報科学研究院
情報学部門

suzuki@inf.kyushu-u.ac.jp

2020年7月21日 (1年用講義)電気情報工学入門I

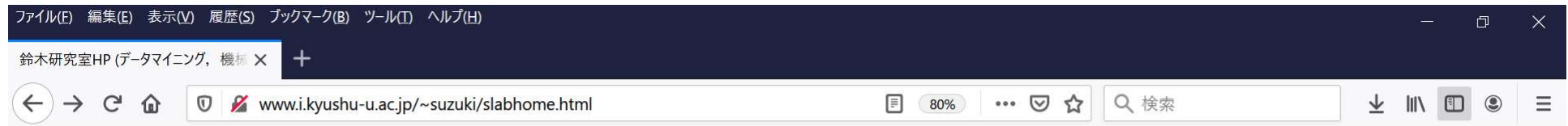


since 1996

鈴木研究室のメンバー

- **現在16名** : 鈴木教授, 松川助教, D学生4名 (全て留学生), M学生7名 (電情B, C課程5名 + 留学生2名), 4年生3名 (電情C課程)
- **修了者と卒業者(延べ)146名(24年間)** : 博士7名, 修士56名, 学士83名
- **最近3年間の就職先** : 日立2名, (国立)東京農工大, 富士通, 富士通研究所, トヨタ, 日産, 日本IBM, 伊藤忠テクノソリューション, 東海理化, Cygames, e成科技, 総務省
- **24年間の就職先** : 次のスライド参照

鈴木研ウェブページ(HP)



鈴木研究室

旧い研究紹介(後日, 今回のスライドを修正してupする予定)

[English version](#)

[データマイニング, 機械学習, 発見ロボット](#)などの研究に取り組んでいます。 [説明スライド](#)。

研究室関連のニュース

(2020年5月) D3の李文博君が[PAKDD 2020@オンライン](#)で発表しました。(当初はシンガポールで開催予定でした。)(2020年2月) 鈴木研と西郷研の歓迎会兼送別会を行いました。M1の波多江 悠祐君が[VISAPP 2020@パレルマ](#)で発表しました。われわれの論文は[最優秀ポスター賞](#)を受賞しました。(2019年11月) M2の大坪悠介君が[ICTAI 2019@ポートランド](#)で発表しました。(2019年4月) 鈴木研と西郷研の歓迎会を行いました。(2019年3月) B4の氏家翔吾君が[甲午賞](#)を受賞しました。M2の本藤拳也君が[火の国情報シンポジウム2019@熊本](#)で発表しました。(2019年2月) 鈴木研と西郷研の送別会を行いました。(2018年9月) 鈴木研の送別会を行いました。(2018年8月) D3の趙開開さんが[ICPR 2018@北京](#)で発表しました。(2018年4月) 鈴木研と西郷研の歓迎会を行いました。(2018年1月) M2の藤田裕文君が[VISAPP 2018@フンメル](#)で発表しました。(2017年12月) 鈴木研と西郷研の歓迎会兼忘年会を行いました。(2017年4月) 鈴木研と西郷研の歓迎会を行いました。 [他のニュース\(構築中\)](#)。

メンバー

教授

[鈴木 英之進](#)

助教

[松川 徹](#)

D3

[李文博](#)

D1

[張 康](#), [鄧 啓鳴](#), [陳 宇](#)

M2

[内村 優太](#), [波多江 悠祐](#), [イサマド フライツ](#), [ウチジマ トモ](#)

M1

[波多江 悠祐](#), [新田 裕章](#), [鈴木 大輔](#), [アン ジュンヒョク](#)

B1

[田中 夏樹](#), [永濱 優基](#), [無津呂 樹](#)

[過去のメンバー\(進路先\)](#)

鈴木の経歴・国際
学会での活動など

研究室の予算で
海外・国内発表など

他の就職先

発表論文(ジャーナル・国際会議など)

発表論文(鈴木研発足以来)

[2021](#), [2020](#), [2019](#), [2018](#), [2017](#), [2016](#), [2015](#), [2014](#), [2013](#), [2012](#), [2011](#), [2010](#), [2009](#), [2008](#), [2007](#), [2006](#), [2005](#), [2004](#), [2003](#), [2002](#), [2001](#), [2000](#), [1999](#), [1998](#), [1997](#), [1996](#)

関連リンク

鈴木は 0. 九州大学, 1. 大学院システム情報科学(研究院, 府) 情報学{部門, 専攻} 知能科学講座, 2. 大学院システム生命科学府 生命情報科学大講座 生命情報発見学講座, 3. 理学部情報科学情報理工学部 電気情報工学科 に所属・教授しています。 [科研費基盤研究\(B\)](#)の研究代表者です。 [JIS](#)の編集委員です。 [ECML/PKDD 2020](#)のエリア長です。 [KDD 2020](#)のプログラム委員会委員です。 [鈴木研究室での訪問者による講演会](#)

トップ論文を審査



ビッグデータに基づくフィンテック

2016/4/21朝日新聞朝刊

5000万円まで貸します(融資).
返済期間6ヶ月, 金利4%台.
電話だけで, 2営業日で振込み.
決算書・事業計画書・担保不要
by アマゾン ジャパン 

アマゾンジャパンへの出店
データを自動
解析.
人手を使わず



新興企業
のため
融資不可



過去の学生からの感想に対する注意

- 企業融資: この条件では破格
- ポイント: アマゾンジャパンが, 出店店舗のビッグデータを収集して自動解析することで, ビジネスで儲けている
- 短く言うと: ビッグデータ × CS技術 × 信用 × 人材 = 大きな利益 (cf. the four GAFA本)
- どんな技術?: 他言しないのが常識
- 学問での意義?: 実運用されている, 確立された技術. 分かりやすく, 学生のやる気を出す

<https://www.amazon.co.jp/four-GAFA-%E5%9B%9B%E9%A8%8E%E5%A3%AB%E3%81%8C%E5%89%B5%E3%82%8A%E5%A4%89%E3%81%88%E3%81%9F%E4%B8%96%E7%95%8C-%E3%82%B9%E3%82%B3%E3%83%83%E3%83%88%E3%83%BB%E3%82%AE%E3%83%A3%E3%83%AD%E3%82%A6%E3%82%A7%E3%82%A4-ebook/dp/B07DNCCG5Q>

ビッグデータ

100Gバイトはビッグデータ？



<https://www.freepik.com/free-photos-vectors/infographic> Infographic vector created by macrovector - www.freepik.com

- Big Data 研究開発 Initiative に2億ドル以上の予算 (オバマ大統領プレスリリース 2012/3/29) 例: 200Tバイト
- 3V (Volume, Velocity, Variety)を満たす数百G [sic], 数Pバイトのデータ (Gartner社)
- ネット通信より, テープを国際郵便で送った方が早い規模のデータ (Jaakko君, 私信)

別の例: 推薦(レコメンデーション)

入力1: 評価行列

抽出情報の例: ユーザ,
アイテムの線形特徴

$$\begin{array}{c} \text{ユーザ} \\ \text{アイテム} \end{array} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \approx \begin{array}{c} \text{ユーザ} \\ \text{アイテム} \end{array} \begin{pmatrix} 1.2 & 0.4 \\ 0.7 & 0.8 \\ 0.3 & 1.0 \end{pmatrix} \begin{array}{c} \text{特徴} \\ \text{アイテム} \end{array} \begin{pmatrix} 0.9 & 0.6 & 0.3 & 0.1 \\ 0.2 & 0.3 & 0.5 & 0.7 \end{pmatrix} \begin{array}{c} \text{特徴} \\ \text{特徴} \end{array}$$

入力2: コンテンツ情報(アイテム情報, ユーザ情報)

出力: ユーザが新規購入するアイテム

昨年度の修士論文のひとつ: ディープラーニングを用いた非線形特徴+ベイズ推論の学習

ディープラーニング(深層学習)

2層ユニットのニューラルネットワーク

$$\mathbf{u} = \mathbf{W}\mathbf{x} + \mathbf{b}$$

$$\mathbf{z} = f(\mathbf{u})$$

非線形関数

入力層

x_1

x_2

x_3

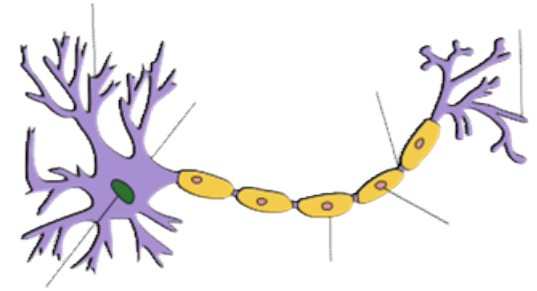
x_4

出力層

$u_1 \rightarrow z_1$

$u_2 \rightarrow z_2$

$u_3 \rightarrow z_3$

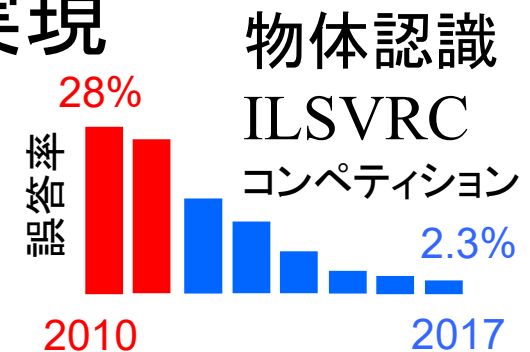


<https://ja.wikipedia.org/wiki/%E7%A5%9E%E7%B5%8C%E7%B4%B0%E8%83%9E>

2018年
チューリング賞

大きな層数で学習すること(深層学習)は困難だったが, Hintonらの長年の努力により実現

種々の分野で大幅な性能向上 + 不可能だった諸タスクを解決



(動)画像説明文(キャプション)生成

訓練フェーズ: 入力された{画像(系列) + {説明文}}
で, ディープニューラルネットワーク(**CNN+RNN**)の
パラメータ群を最適化

畳み込みNN

テストフェーズ: 入力された新規画像(系列)の説明文を生成



a **man** playing the **piano** in the **park**

深層学習以前: a man sitting on the chair



a woman is cutting

[Oura et al. IJCNN 18] (**修士研究**): 動画画像説明文生成での静止画用NNの利用

ワンショット画像生成

入力は

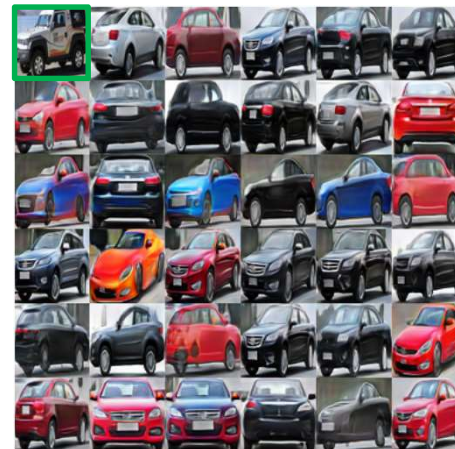
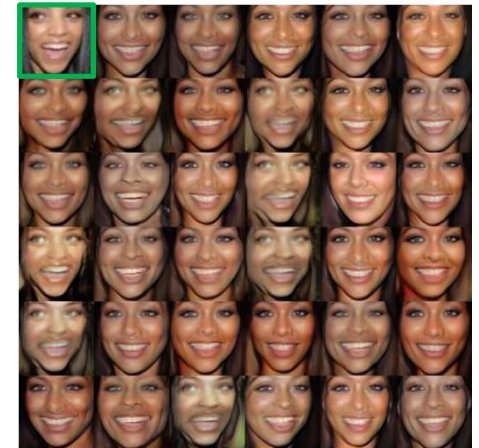
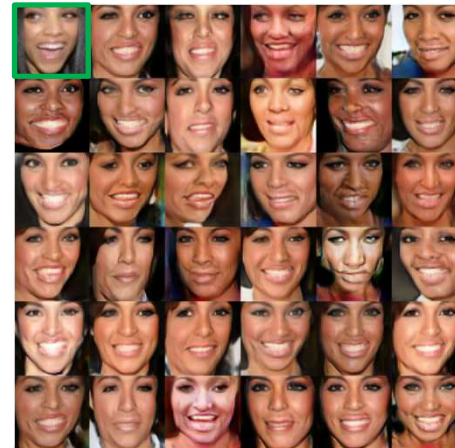
左上の1枚

訓練フェーズ: 入力された
{画像 + クラスラベル} で,
ディープニューラルネット
ワーク(**GAN: 敵対的生
成ネットワーク**)のパラ
メータ群を最適化

テストフェーズ: 入力され
た新規クラスの画像1枚
から, 同じクラスの画像群
を生成

DAGAN [Antoniou 17]

Tri-DAGAN

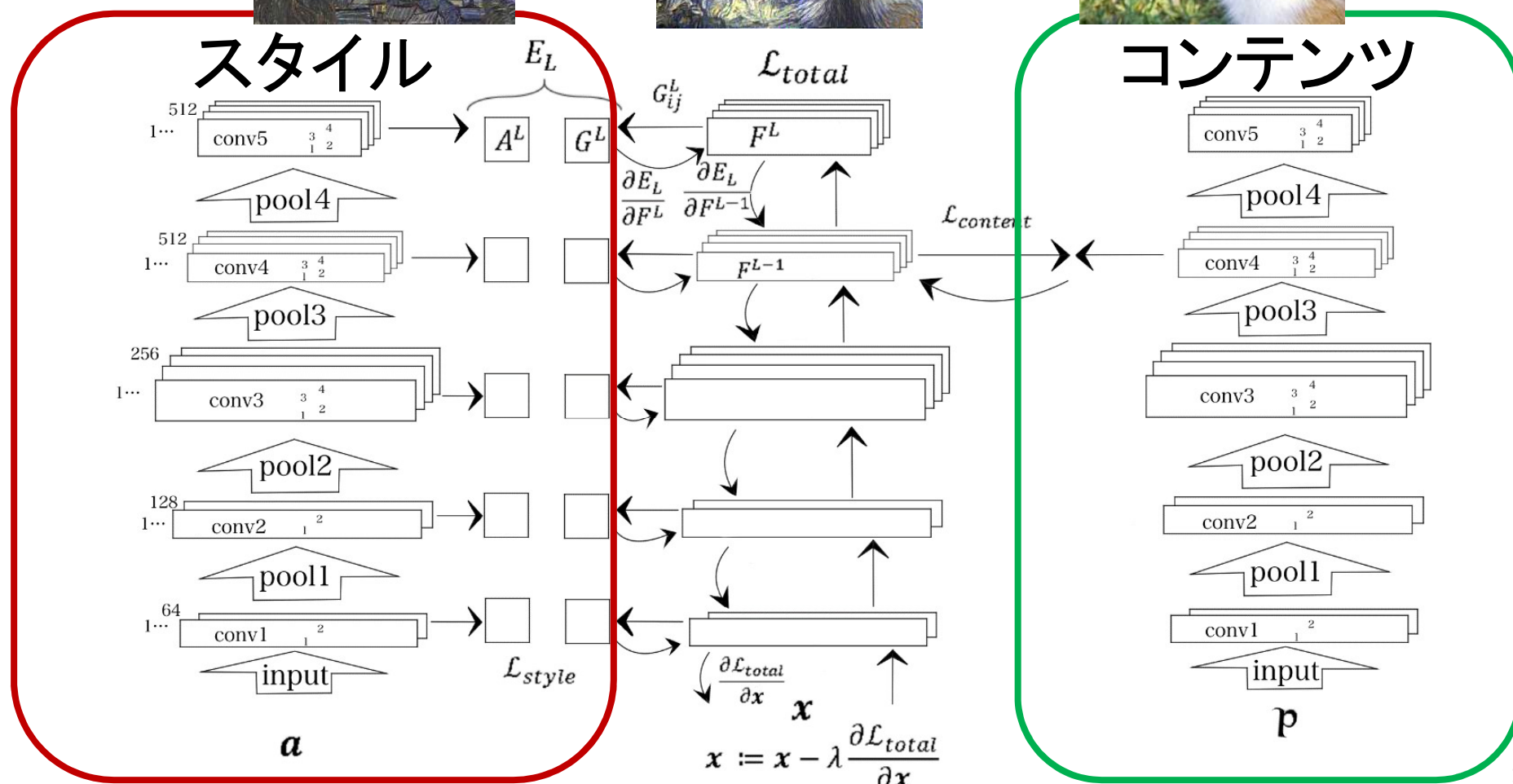
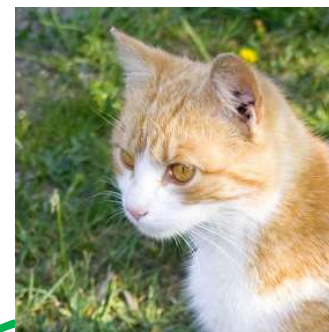


[Ohtsubo et al. ICTAI 19] (**卒業研究**): 細粒度カテゴリ
データが対象. 距離学習との統合

左右の
画像から中央
の画像
を生成

スタイル変換 [Gatys16]

未発表
の卒業
研究
あり



学生たちの研究テーマ種類の例

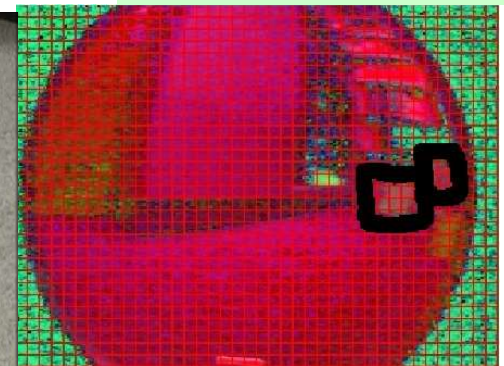
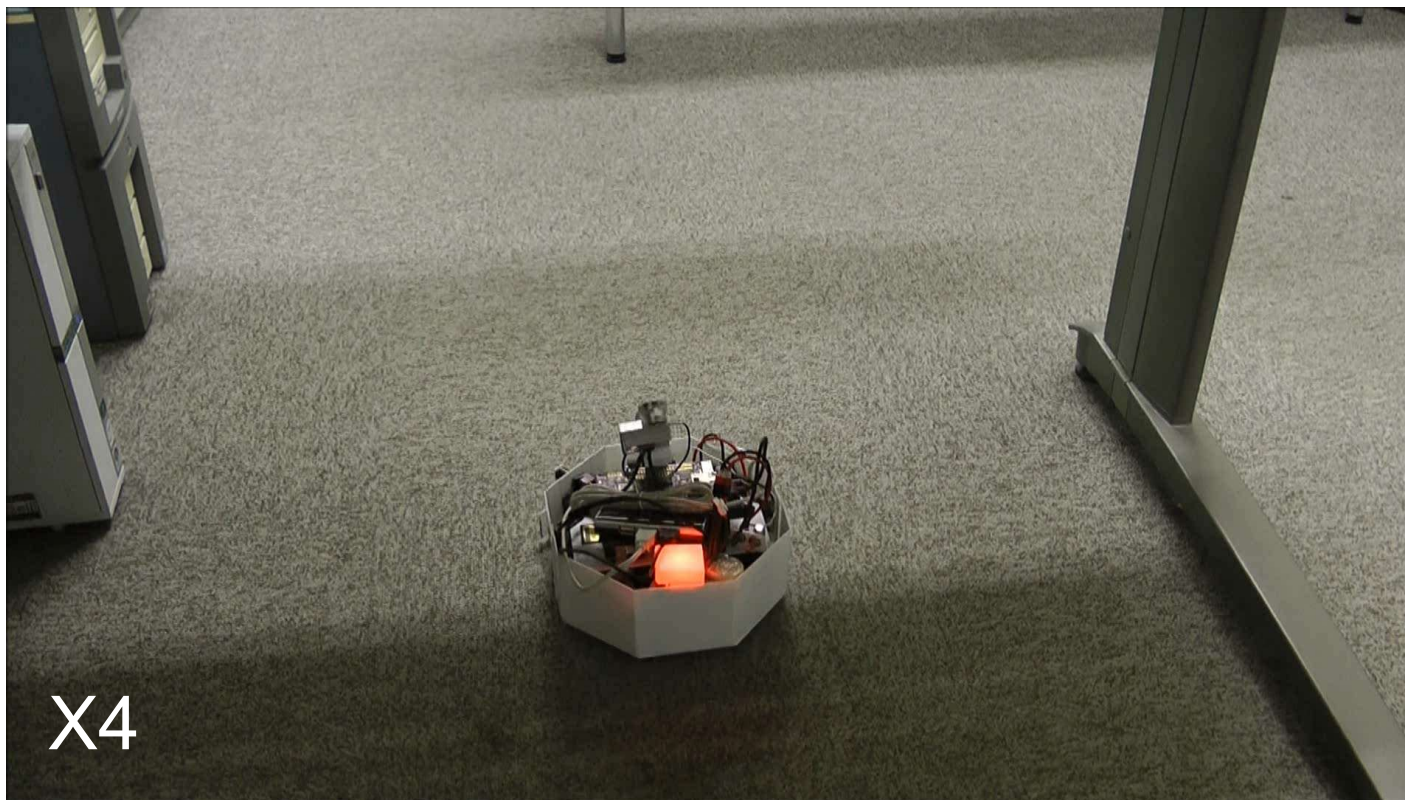
強制・限定せず，相談して数か月間で決定（考え，サーベイし，予備実験を行うことが必須）

- （常に）機械学習・データマイニング手法（KDD国際会議，ICDM: IEEEデータマイニング国際会議，ICML: 機械学習国際会議で論文発表や役職を務めた立場からアドバイス）
- （主に助教授時代）実データからのデータマイニング（手法ではなく地頭. 新規手法の論文発表とセット）
- （最近人気）ディープラーニング（就活と就職後にも関係）
- （最近6年）顔表情の学習（約半数の人は，感情が現れる）
- （最近10年）自律移動ロボットによる学習と発見

自律移動ロボットによる異常検知 [Suzuki 2012]

ロボットがデータの計測(この例は10分間. 1枚/秒. 120万部分画像), 色のモデリング(色のbagで**CF木** [Zhang 96]を構築), 異常値の検知(最後の9分間)を行う

高さ平衡木



転倒リスク発見（修士研究・博士研究, 国際ジャーナルや国際会議で多数発表. 2013-2017）

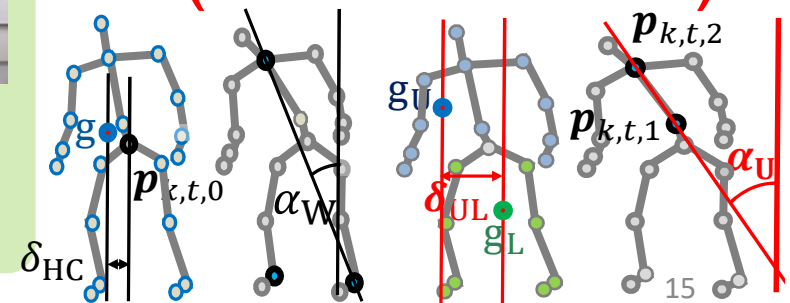
例はKinectで計測した骨格. クラスタリング



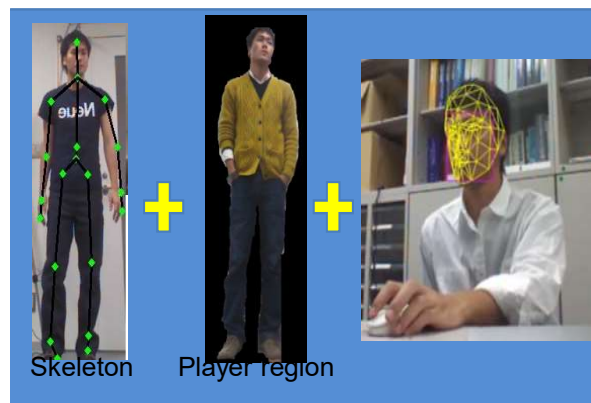
発見されたリスク姿勢（クラスタ）



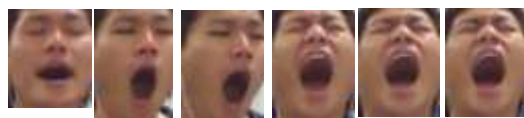
仏との共同プロジェクト(JSPS-CNRS)



顔表情クラスタリング (修士研究 [Kondo et al. Aml 14])

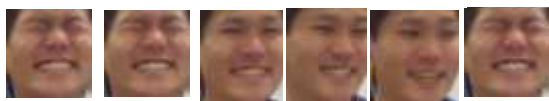


クラスタ 1



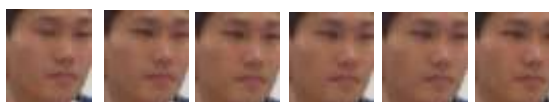
顔表情は感情 (fast思考) と連動

クラスタ 2



顔の表情筋に関する特徴6個

クラスタ 3



+ピッチ角を用いた新規特徴

説明文つき画像系列からの異常検知 (卒業研究 [Hatae et al. VISAPP 20])

正常例だけ

Best poster award (本件のみ)

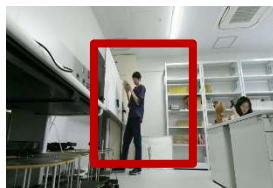
異常例あり



“woman standing in front of a white wall”



“white cabinet with white door”



“man standing in front of a door”

キャプションは類似・同一
ただし**不正確**

訓練
フェーズ

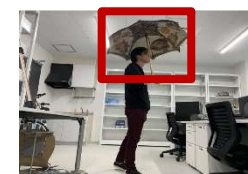
テスト
フェーズ



画像特徴, 説明文特徴,
位置特徴



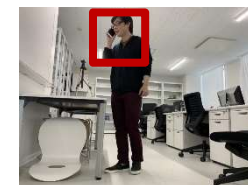
“woman standing in front of a white wall”



“a large umbrella”



“a stack of books on the bed”



“man holding a cell phone”