[共通セッション] 土木分野におけるAIの活用

2024年9月5日(木) 13:00 ~ 14:20 第2講義室(経済学部)(川内南キャンパス文科系総合講義棟)

土木分野におけるAIの活用（その3）

座長：伊東 広敏（エイト日本技術開発）

13:30 ~ 13:40

[CS14-18] マルチモーダルモデルを用いたアクティビティ調査手法の提案

*藤井 純一郎[1]、岡野 将大[1]、吉田 龍人[1]、高森 秀司[1] (1. 八千代エンジニヤリング株式会社)

キーワード：ウォーカブル、公共空間、アクティビティ調査、マルチモーダルモデル、深層学習

国土交通省は，人中心で居心地が良く歩きたくなる空間をまちなかに創出することを，新たな都市政策として掲げている．しかし，居心地の良さを評価する手法は人手での作業に大きく依存するため，膨大な労力を要することが問題となっている．
本研究では，公共空間の居心地の良さに関連するアクティビティ調査を対象として，マルチモーダルモデルを用いて自動化する手法を提案した．提案手法の妥当性確認のため，基礎的な実験を行った．

© 2024 Japan Society of Civil Engineers

マルチモーダルモデルを用いたアクティビティ調査手法の提案

八千代エンジニヤリング㈱　正会員　○藤井 純一郎，吉田　龍人，高森　秀司
八千代エンジニヤリング㈱　非会員　岡野 将大

１．背景と目的

国土交通省は，まちなかでの人中心で居心地が良く歩きたくなる空間の創出を新たな都市政策として掲げている[1]．並行して公共空間の居心地の現状把握や改善効果を計測するための評価指標も策定されている[2]．一方で，評価指標の算出方法は，現地での人手によるカウントや撮影動画の目視カウントが主流となっている．いずれの方法も人手での作業に大きく依存しており，その自動化はウォーカブル関連施策の推進においても重要な意義がある．

都市空間の賑わいを評価するうえでは，滞在者数等の量的な把握に加えて活動内容等の質的な評価が重要となる．量的評価の自動化は画像認識技術の適用が進みつつあるが，質的な評価の自動化はいまだ途上にあると言ってよい．

そこで，本研究では公共空間の質的評価のための調査であるアクティビティ調査を対象として，マルチモーダルモデルを用いて自動化する手法を提案する．また，提案手法の妥当性を確認するため，基礎的な実験を行い，その精度を定量評価する．

２．マルチモーダルモデル

従来のAIモデルは，言語処理，画像認識，音声認識など，特定のデータを対象としてきた．これに対して，複数のデータ種別を併用して問題を解くマルチモーダルモデルと呼ばれる技術が進展している．

特に言語と画像の関係性を学習したVision & Language（V&L）モデルの進展が著しく，大量のデータで事前学習したV&Lモデルは，追加の学習を行うことなく様々なタスクに適用できる汎用性を持つ．本研究ではVision Question Answering（VQA）タスクと呼ばれる，V&Lモデルで画像に対して質問回答を行うタスクをアクティビティ調査に適用する．

本研究では，オープンソースで事前学習済みモデルが公開されており，VQAに対応しているLLaVA1.5 Vicuna1.5-13B[3]を用いた．また，モデルの都合上，各質問のプロンプトは英語とした．

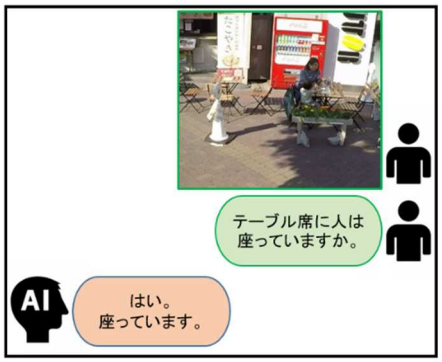

図-1　ﾏﾙﾁﾓｰﾀﾞﾙﾓﾃﾞﾙによる VQA のｲﾒｰｼﾞ

３．提案手法

公共空間のアクティビティ調査では，一般的に複数の行動種別を記録することが求められる．本研究では，画像から複数の行動認識をするために，マルチモーダルモデルを用いたVQAを複数回実施する．VQAの質問内容は，画像内の行動を分類するため，二項選択式の問いとし，その回答を集計する．

図-2に本研究で提案するアクティビティ調査手法のフローを示す．はじめに，画像内に調査対象となる人物がいるかYes/Noの二値分類で判定する．次に，調査対象の施設（本研究ではベンチ）の使用の有無を同様に判定し，回答が「Yes」の場合，ベンチ利用者のアクティビティに関する質問を行う．

さらに，長時間の動画に対して調査を行うため，動画を一定時間間隔の画像に分割するフレーム抽出を行った上で，各フレームに対して連続的にVQAを適用し，図-3に示す集計を行うことで，従来の目視調査と同等の時間単位の記録を可能とした．

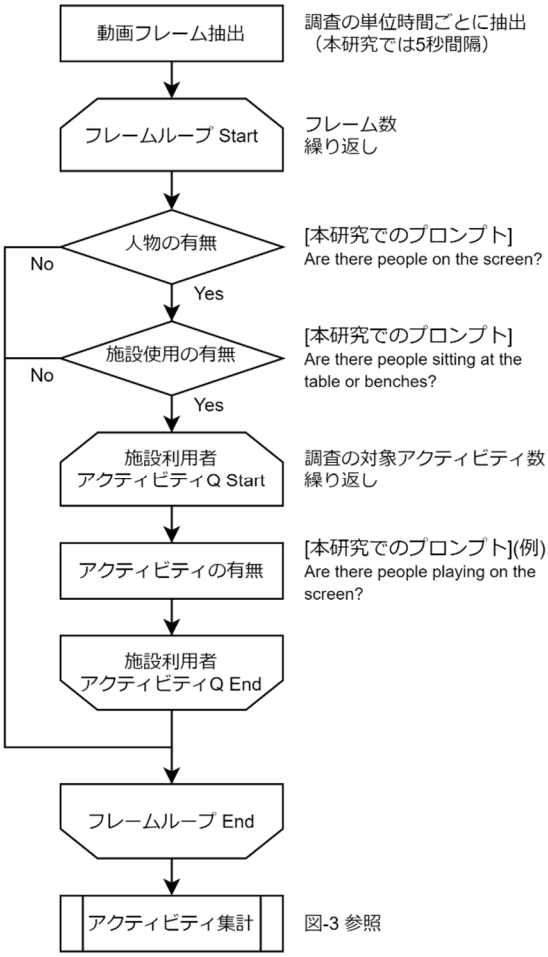

図-2　提案手法フロー図

キーワード　ウォーカブル，公共空間，アクティビティ調査，マルチモーダルモデル，深層学習
連絡先　〒111-8648　東京都台東区浅草橋 5-20-8 CS タワー　八千代エンジニヤリング㈱　TEL:03-5822-6626

© Japan Society of Civil Engineers

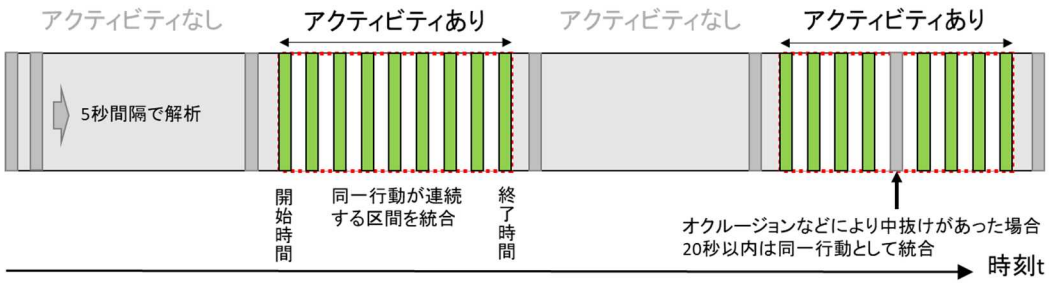

図-3　アクティビティ集計のイメージ

表-1　調査結果の例（○：アクティビティあり）

ベンチ利用開始	ベンチ利用終了	ベンチ利用時間	飲食	談話	携帯	休憩
2:15	14:45	12:30	○	○	—	—
4:10	4:35	0:25	—	—	—	○
6:40	8:50	2:10	—	—	○	—
12:25	17:20	04:55	—	—	○	—
12:35	17:20	04:45	—	—	—	○

表-2　精度評価結果

Activity	Accuracy 正解率	Precision 適合率	Recall 再現率	Specificity 特異性
ベンチ利用	0.786	0.955	0.724	0.923
飲食	0.881	1.000	0.688	1.000
談話	0.762	0.692	0.600	0.852
携帯	0.952	0.600	1.000	0.949
休憩	0.810	0.400	0.286	0.914

4．実験

4.1. 実験概要

提案手法を既存のアクティビティ調査で取得した動画に適用した．表-1に示す既存の調査で目視により抽出したベンチ利用に関するアクティビティ記録を正解とし，提案手法で自動取得した記録と比較した．対象とするアクティビティは，ベンチ利用に加えて，ベンチ利用者が何をしていたかを{飲食/談話/携帯電話利用/休憩のみ}に分類したものである．

4.2. 実験データ

実験には，3箇所のベンチの利用状況を撮影した16本の動画を用いた．各動画は20分あり，30FPSで撮影されている．これらの動画から5秒間隔のフレームを抽出し，トータルで3,840枚の入力画像を得た．画像をモデルに入力する際は，LLaVA-1.5の入力サイズである336px×336pxにリサイズした．

4.3. 実験結果と考察

評価指標として，目視調査に対する正解率のAccuracy，モデルでアクティビティありとした記録の正しさを評価するPrecision，目視調査でアクティビティありとした記録の見落としの少なさを評価するRecallを算出した．加えて本実験ではアクティビティなしの時間帯が長いことから，それを評価するSpecificityも算出した．評価結果を表-2に示す．

最も基本的なアクティビティであるベンチ利用のAccuracyは78.6%であった．すなわち，人手による目視調査に対して，2割強ほどの記録で間違いが発生している．間違いの内容は，Precision，Recall，Specificityを見るとRecallが低い結果であったため，提案手法でアクティビティの見落としが発生していることが分かる．見落としがあった記録部分の動画を確認すると，目視調査では複数のアクティビティに分割して記録されていたものが，提案手法では統合されて1記録となっていることが原因であった．これは，VQAタスクでは正しくアクティビティが認識できており，それを時間単位で集計する際の間違いであるため，集計ロジックを改善することで精度向上が可能と考える．ただし，正解とした目視調査結果では複数の記録に分かれていたものの，1記録として統合してもおかしくない記録も見られたため，必ずしも提案手法の結果が間違いと判断できない．

その他のアクティビティも，ベンチ利用とほぼ同等かそれ以上のAccuracyを示した．一方で，Precision/Recallの値は大きくばらつく結果であった．これは，各アクティビティの記録数が少ないため，少量のFP/FNがあるだけで値が大きく変動するためと考察した．精度を正しく評価するには，より長時間の調査に適用する必要がある．

5．まとめと今後の課題

本研究では，マルチモーダルモデルを使用したアクティビティ調査の自動化手法を提案した．提案手法の妥当性確認のため，基礎的な実験を行った結果，提案手法は目視調査に対して，約8割の正解率であった．精度向上に向けては，時間単位のアクティビティ集計方法の改善が課題である．また，多地点・長時間の調査で精度検証を行うことも課題である．

提案手法の実用化に向けては，調査時間や作業量の省力化についても検証を行いたい．アクティビティ調査が省力化することで，公共空間の質的評価の高密度化・高頻度化とデータの蓄積が実現できる．データに基づく施策の検討と改善が図られることで，豊かな公共空間形成に資することを期待する．

参考文献

1) 国土交通省: ストリートデザインガイドライン ver2.0, 2021.
2) 国土交通省: まちなかの居心地の良さを測る指標改訂版 ver1.0, 2023.
3) Haotian Liu, Chunyuan Li, Yuheng Li, Yong Jae Lee : Improved Baselines with Visual Instruction Tuning, arXiv:2310.03744, 2023.

© Japan Society of Civil Engineers