

平成30年度 修士論文

レスキュー犬の一人称動画
を用いた動作分類

電気通信大学大学院 情報理工学研究科

情報学専攻 メディア情報学コース

1730010 荒木 勇人

指導教員 柳井 啓司 教授

平成31年1月30日

概要

被災地での災害救助を補助する犬をレスキュー（災害救助）犬といい、カメラなどの計測装置を装備したレスキュー犬をサイバーレスキュー犬と言う。本研究では、犬にとりつけたセンサからサイバーレスキュー犬の活動を識別した。

目次

第1章 はじめに	1
第2章 関連研究	3
第3章 提案手法	5
3.1 動画像平均画像クラス分類	5
3.2 オプティカルフロー画像クラス分類	5
3.3 音声クラス分類	5
3.4 音声と動画のマルチモーダル情報クラス分類	5
第4章 データセット	6
4.1 DogCentric Activity Dataset (DCAD)	6
4.2 サイバーレスキュー犬 訓練データセット	7
第5章 実験結果	8
5.1 シングルクラス分類	9
5.1.1 動画像平均画像クラス認識	9
5.1.2 オプティカルフロー画像クラス認識	9
5.1.3 音声クラス認識	9
5.1.4 音声と動画のマルチモーダル情報クラス認識	9
5.2 マルチクラス認識	9
5.2.1 動画像平均画像クラス認識	9
5.2.2 オプティカルフロー画像クラス認識	9
5.2.3 音声クラス認識	9
5.2.4 音声と動画のマルチモーダル情報クラス認識	9
第6章 まとめ, 今後の課題	10

第1章

はじめに

被災地での救助活動を行う際に、訓練されたレスキュー犬（災害救助犬）が人間の補助として探査を行う場合がある 図1.1。災害救助犬を育成し、現場に派遣する団体は日本国内に複数存在し、必要に応じて現場に派遣される。レスキュー犬は、犬としての特性を生かして人間と協力して被災地の探索を行う。レスキュー犬にはがれきの隙間などの狭い空間、倒壊した建物など人間には踏破困難な環境でも探査可能であったり、またその発達した嗅覚を頼りにした探査が可能である。このように、人間では探査が困難あるいは不可能な環境においても人間の能力をレスキュー犬が補うことで効果的な救助活動が期待される。しかし、彼らは人間に向けた言語を持たない。そのため、人間はレスキュー犬の行動をよく観察し、彼らが収集した情報を彼らの様子から推察、理解しなくてはならない。現状では、レスキュー犬を指揮するハンドラーと呼ばれる人間がレスキュー犬の行動を手動でマーキングして犬の周辺環境の情報収集と理解に努めている。収集された情報は消防などのハンドラーらを統括する指揮命令者に口頭伝達され、現場の把握に活かされる。このレスキュー犬と人間との共同探索の問題点として、トリアージ（緊急度に従った手当の優先順位付け）のための災害現場周辺環境情報や、要救助者情報の不足があげられる。また、ハンドラーによる記録はどうしても主観的になり客觀性が不足し、さらにそれを口頭伝達することで正確性がより不足する。レスキュー犬によって収集された情報を個人の主觀に基づくことなく分類し、整理された情報を共有できれば災害救助活動の効率化がより図れる。

本研究では、レスキュー犬にセンサを装着して得られたデータ用いてレスキュー犬の行動をリアルタイムに分類すること目的とする。深層学習を用いた画像識別にある既存手法を予備実験として行った。予備実験をもとに、動画からのレスキュー犬行動分類を行う。本研究は映像だけでなく音声などのデータも活用したマルチ

モーダルな動画分類である。本研究により、レスキュー犬が今何をしているのか明示的に判断することが可能となり、トリアージに必要な情報が整理され、災害救助活動の効率化が期待される。



図 1.1: 被災地におけるレスキュー犬らの救助活動 [?] より引用

第2章

関連研究

レスキュー犬の行動をモニタリングするために、濱田、大野らによって装着型計測・記録装置が開発された [?]. 図2.1にレスキュー犬に装着可能な軽量な行動計測スーツを示す。これを着用したレスキュー犬はサイバー救助犬とも呼ばれる。各種センサを用いた計測データを記録し、リアルタイムに映像などのデータを無線配信することが可能である。そのため、レスキュー犬が人の目の及ばない範囲で活動する際にもレスキュー犬の行動やその周辺環境などを把握するのに役立つ。サイバー救助犬は、政府による総合科学技術・イノベーション会議が研究開発を促進しているImPACTというプログラムのタフ・ロボティクス・チャレンジの一環である。タフ・ロボティクス・チャレンジとは災害救助を目的としたロボットの研究開発プロジェクトであり、その中で、災害救助用サイボーグ犬の開発の足掛かりとしてサイバー救助犬が研究されている。



図2.1: 装着型計測・記録装置 [?] より引用

また、Ehsanらによる犬の一人称視点動画からの犬行動予測の研究がある [?].

これは、犬の行動をモデリングし、犬が次にどのような道をたどり行動するかを予測している。

しかし、これらの研究は犬の行動のモデリングであり、犬の周辺環境の推定などは行っていない。また、入力は動画像のみであり、音声などのデータは利用していない。レスキュー犬の課題には、犬の周辺環境情報や動画像からだけでは判断できない情報の取得が含まれている。例えばレスキュー犬は要救助者を発見するとその場で待機し吠え続けるように訓練されている。このように、動画像データからだけではなく、音声データ、および慣性データ・GPSデータなどの情報を複合的に用いてレスキュー犬の状態を判断しなければならない。本研究は動画像と音声からなるマルチモーダルな情報を入力とした犬の行動の分類を目的としている。

第3章

提案手法

単一クラス分類とマルチクラス分類をそれぞれ行った。犬一人称視点動画单一クラス分類では、

- 3.1 動画像平均画像クラス分類
- 3.2 オプティカルフロー画像クラス分類
- 3.3 音声クラス分類
- 3.4 音声と動画のマルチモーダル情報クラス分類

第4章 データセット

既存の公開されている犬一人称視点動画データセットに DogCentric Activity Dataset(DCAD) がある。本研究ではレスキュー犬向けにラベル付けされたレスキュー犬訓練動画が必要であるため、本実験ではレスキュー犬の訓練動画を用いた。訓練動画を用いる前に、犬一人称視点動画から行動分類が可能かどうかを確認する予備実験を行なった。予備実験には DCAD を用いた。

4.1 DogCentric Activity Dataset (DCAD)

4頭の犬の背中に GoPro カメラを取り付けて散歩をした動画を单一クラス分けしたデータセット。動画は 320 x 240 解像度、48 frames per second で撮影されている。散歩する地域やコースは犬毎に異なり、アノテーションはそれぞれの犬に同じラベルのアクティビティをラベル付けしている。アクティビティは 10 クラス（横断前の待機: Car、水分の摂取: Drink、手渡しでの食事: Feed、左を向く: Look at left、右を向く: Look at right、人間が犬を撫でる: Pet、ボールで遊ぶ: Play with ball、身体をブルブルと振る: Shake、何かの匂いを嗅ぐ: Sniff、歩く: Walk）あり、それぞれ合わせて 209 クリップになる 4.1.

表 4.1: DogCentric Activity Dataset 内訳

Activity	Car	Drink	Feed	Left	Right	Pet	Ball	Shake	Sniff	Walk
Clips	26	10	25	21	17	25	14	19	27	25

4.2 サイバーレスキュー犬 訓練データセット

第 5 章

実験結果

予備実験の結果を (図 5.1,5.2) にそれぞれ示す。分類率は、VGG16 モデルを利用したものが 64.3%, ResNet モデルを利用したものが 59.5% であった。全般的に、データの多いクラスは精度が高い傾向にあるが、データの少ないクラスは精度が低い傾向にある。加えて、Car クラスは道路の進行方向に対して垂直に待機している 10 クラスの中で特殊なクラスであり、車などの写ったフレームの影響で分類精度が上昇していると考えられる。Feed クラス, Pet クラス, Play_with_ball クラスは、それぞれフレーム内を人間が占める割合が多いクラスと言え、そのため混同が起こりやすいと考えられる。

	Car	Drink	Feed	Left	Right	Pet	Ball	Shake	Sniff	Walk
Car	6	0	0	0	0	0	0	0	0	0
Drink	0	1	0	0	0	0	0	0	2	1
Feed	0	0	1	0	0	0	0	0	0	0
Left	1	0	0	1	0	0	0	0	2	0
Right	0	0	0	0	0	0	0	0	1	0
Pet	0	0	1	0	0	3	1	0	2	0
Ball	0	0	0	0	0	0	5	0	0	0
Shake	0	0	0	0	0	0	1	1	2	0
Sniff	0	0	0	0	0	0	1	0	3	0
Walk	0	0	0	0	0	0	0	0	0	6

図 5.1: VGG16 pretrained model による finetuning の結果

	Car	Drink	Feed	Left	Right	Pet	Ball	Shake	Sniff	Walk
Car	6	0	0	0	0	1	0	0	0	0
Drink	0	1	0	0	0	0	0	0	0	0
Feed	0	0	1	0	1	1	1	0	0	0
Left	0	0	0	1	1	0	0	1	0	2
Right	0	0	0	1	2	0	0	0	0	0
Pet	0	0	2	0	0	3	1	2	1	0
Ball	0	0	0	0	0	0	2	0	0	0
Shake	0	0	0	0	0	0	0	1	1	0
Sniff	0	0	0	0	0	0	1	0	5	0
Walk	0	0	0	0	0	0	0	0	0	3

図 5.2: ResNet pretrained model による finetuning の結果

5.1 シングルクラス分類

5.1.1 動画像平均画像クラス認識

5.1.2 オプティカルフロー画像クラス認識

5.1.3 音声クラス認識

5.1.4 音声と動画のマルチモーダル情報クラス認識

5.2 マルチクラス認識

5.2.1 動画像平均画像クラス認識

5.2.2 オプティカルフロー画像クラス認識

5.2.3 音声クラス認識

5.2.4 音声と動画のマルチモーダル情報クラス認識

第6章

まとめ,今後の課題

動画の各フレームの平均を取り, 画像として識別した。データの少ないクラスは精度が低いため, データを補う必要がある。予備実験では簡易的な方法を用いたが, 今後は最新手法による分類を検討している。またレスキュー犬の行動を認識する際には複数クラスの出力にする必要がある。

今後の課題として, 時系列情報を特徴量抽出に使う。また, 音声データから特徴量を抽出し, 動画特徴量と併せたマルチモーダルな特徴量を利用し, レスキュー犬の行動分類を行う。