

RACRA: 定点カメラ映像用高速アノテーションツール

白井 良成^{1,a)} 水谷 伸¹ 岸野 泰恵¹ 松岡 舞² 北川 忠生² 納谷 太¹

概要: 定点カメラで撮影した映像に対するラベル付けを高速に実現可能なアノテーションツール, RACRA(Rapid Annotation Creation for Rapid Analysis) を提案する. RACRA は映像中の複数のセグメントに存在する動物体を重畳表示する. ユーザは, この重畳映像を利用して, 動物体へのラベル付けを切れ目なく次々に行うことができる. 我々は本ツールを, 生物の行動を詳細に理解するために撮影した映像へのラベル付けを目的としてデザインした. このような映像へのラベル付けは, 生物の微細な差異に関する知識を必要とするため画像処理による自動化が難しく, また, ラベル付けできる人も限られており, 大量の映像データの分析は困難であった. RACRA により高速なラベル付けを実現することで, 大量の映像データに基づく生物行動理解を促進できると考えている.

RACRA: Rapid Video Annotation Tool for Fixed-Point Camera Images

YOSHINARI SHIRAI^{1,a)} SHIN MIZUTANI¹ YASUE KISHINO¹ MAI MATSUOKA² TADAO KITAGAWA²
FUTOSHI NAYA¹

Abstract: We have developed a novel video annotation tool called RACRA. RACRA enables users to rapidly annotate labels into the fixed-point camera images. Overlaying multiple moving objects belonging to different segments of the video allows users to keep track of multiple labeling targets at all times. Thus, users can continuously annotate labels on the video. We have conducted a test using underwater camera images to improve the RACRA. This paper describes the design and implementation of RACRA, and discusses the effectiveness and limitations of our approach.

1. はじめに

動物行動学において, 情報科学的な手法を活用して行動の特徴や群れの社会構造に迫る “computational ethology” が現れつつある [1], [4]. computational ethology では, 映像中における動物の特定行動数のカウント, 動物の軌跡などのデータを動画解析により取得し, そのデータを用いて動物の行動の定量化を行う [4]. 動画の撮影と解析を行うことで, 映像を用いなかった場合に比べ, 様々な行動のパラメータを定量化することができ, 動物の行動に関する客観的かつ多様なデータに基づいた動物行動の理解が可能となる. このような動画解析による行動分析を支援するツ

ルもすでにいくつか公開されている [5], [10].

しかしながら, より高度な分析を実施しようとする, 既存の画像処理技術では, その行動を十分に抽出できないという課題がある. 例えば, 自然環境下で撮影された映像では, 分析のターゲットとする生物以外にも様々な生物が映り込んでしまう. 個体や行動の分類において, 色や形, 動きなどを総合的に判断しなければ分類できないような識別は, 現在の画像処理技術ではまだまだ困難であり, 人手によるアノテーションが必要となる.

そこで我々は, 画像処理技術とインタラクション技術を融合させた, 映像解析ツール: RACRA: Rapid Annotation Creation for Rapid Analysis) を作成している. RACRA は, 画像処理技術によって抽出された映像中の動物体に対する, 高速なアノテーション付けを支援する.

本論では, RACRA デザインと実装について述べ, RACRA の可能性について議論する.

¹ 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
NTT Communication Science Labs, NTT Corporation

² 近畿大学 農学部 環境管理学科
Department of Environmental Management, Faculty of Agriculture, Kindai University

a) shirai.yoshinari@lab.ntt.co.jp

2. アノテーションを利用した動物行動分析

2.1 アノテーションツール

談話解析の研究領域では、ELAN 等の映像、音声に対するアノテーションを付与するツールが利用されている [12]. また、画像処理の研究領域でも、Grand Truth データの作成や映像の利用促進を目的として、映像に対するアノテーション（ラベル付け）ツールが提案、構築されている [2], [7], [11]. 手作業によるラベル付けは、正確性が高い反面、ビデオの量が増えれば増えるほど、膨大な時間が必要となる. そこで、いくつかのツールでは、アノテーションにかかる時間と手間を減らすため、画像処理技術を利用した半自動アノテーション [2] やクラウドソーシング [11] を用いた負荷分散など機能を実現している.

しかしながら、これらのツールは、computational ethology に利用するうえで、以下の課題がある.

(1) アノテーションに必要な知識

高度な行動理解を行うためには、動物やその行動、軌跡に対するラベル付けに、専門的な知識が必要となる. 専門的知識を持つ人は限られるため、クラウドソーシングなどのアプローチは往々にして利用が難しい.

(2) アノテーションに要する時間

高度な行動理解を行うために必要な、個体や行動の識別（ラベル付け）を自動化することは現在の画像認識技術では困難である. そのため、専門的な知識をもった人が手作業で行うこととなるが、映像に対するラベル付け作業は通常映像そのものの時間以上の時間を要するため、長時間にわたる映像に対するラベル付けを専門知識を持った少数の人で行うのは困難である.

専門知識が必要なわかりやすい事例として雌雄判別が挙げられる. 一見してその個体が雄か雌かの判断が難しい生物は多い. 魚であれば、例えばヒレの形などが判断材料となるが、それらを動画中から素早くかつ正確に判別することは、ある程度の知識を持った人でなければ困難である. 同様に同科、同族内の生物の分類なども素人には往々にして困難であろう.

2.2 希少魚生息池における映像の記録

映像を利用した動物行動解析の具体例として、希少魚生息池モニタリングの取り組みを紹介する. 我々は現在、絶滅危惧種の魚、ニッポンバラタナゴの産卵行動を理解するため、センサネットワークを用いた環境モニタリングを実施している [6], [9]. 2015 年からは収集した水温や溶存酸素量などの測定データと産卵行動との関係を明らかにするため、同タナゴが生息する池に水中カメラを設置し、タナゴが産卵する貝をターゲットに映像の長時間録画を開始している [8]. 映像から貝にアプローチするタナゴの雌雄別個体数等の推移を確認し、水温や溶存酸素量との関係を明

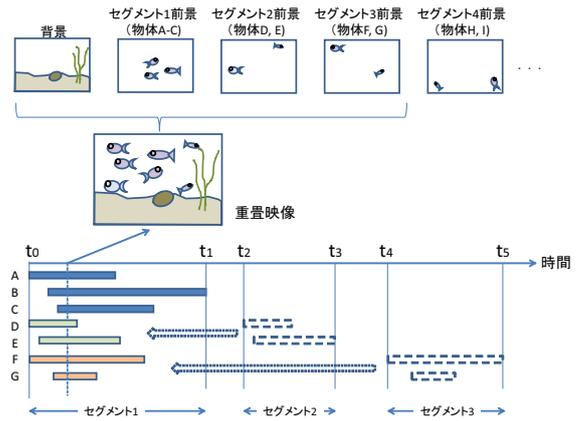


図 1 重畳映像の作成

Fig. 1 Overlaying multiple segments of the video

らかにすることが狙っている. しかしながら、池にはタナゴ以外にもミナミメダカなど数種の魚が生息しており、映像中に移り込んだ魚がタナゴか否か、また、それが雄か雌かといった識別は、現在の画像認識技術ではまだまだ困難であり、かつ、ある程度日常的に魚を観察しているものでなければ難しい.

3. RACRA: Rapid Annotation Creation for Rapid Analysis

専門的知識を持った人が高速にラベル付けできることをめざし、我々は定点カメラで撮影された映像に対する高速アノテーションツール、RACRA を現在構築中である. RACRA のデザイン及び実装について以下に述べる.

3.1 視覚情報処理に着目したインタラクションデザイン

RACRA は人間の視覚が並列処理可能なこと [3] に着目しデザインを行った. RACRA は、まず、背景差分やオブジェクトトラッキングなどの画像処理を用いて、映像中の動物体を背景から切り出し、軌跡データを生成する. 次に、抽出された軌跡データ群を基に、軌跡データが存在する時区間と、存在しない時区間にセグメンテーションする. そして、あらかじめ指定したセグメント数に基づいて、複数のセグメントにおける軌跡映像（背景から分離した静止画像群）を、背景画像に重畳して表示する.

3つのセグメントの重畳例を図1に示す. 異なる時区間の物体 A-G が背景画像と共に重畳され、 t_0 から t_5 の時区間が t_0 から t_1 の映像として合成される. ユーザは、この異なる時間帯における動物体が重畳された映像に対して、ラベル付けしたい個体をクリックする等の方法により、ラベルを付与していく. ラベル付けされた軌跡映像は合成映像中から消滅する. 特定のセグメントに属する軌跡映像すべてにラベル付けが終了すると、表示セグメント数を維持するために新たな（時間的に後の）セグメントが重畳され

ていく。図1においては、例えば、物体 D,E のラベル付けが終了すると、セグメント 2 のラベル付けがすべて終了したと判定されて、新たにセグメント 4 の物体 H, I が背景を除いた最奥に重畳される。これにより、ユーザは、絶えず複数のラベルリングターゲットを見ながら、高速にアノテーションの付与をしていくことが可能となる。

3.2 RACRA の表現と操作

希少魚生息池において撮影した水中カメラ映像を読み込んだ RACRA のスクリーンショットを図 2 に示す。本スクリーンショットでは、時間の異なる 10 のセグメントに属するタナゴやミナミメダカなどの動物体が合成されている。図 1 下部と同様の表現で、個別の動物体の出現時間がそれぞれタイムラインビューが表示される。同一セグメントに属する動物体は、同色の矩形枠で囲われる。これにより、動物体同士の関係（雄が雌に接近中等）のコンテキストは維持される。ユーザは、タイムライン下部のパネル上でラベルを指定し特定の矩形をクリックすることで、その動物体に“たなご 雄”，“みなみめだか 雌”といったラベルを素早く付与していくことができる。

現在、ラベル付け操作としては、上記マウスによるクリックのほか、キーボードショートカットによる操作を実現している。重畳される動物体は、現在、映像の先頭に近いものほど、重ね順の先頭に表示されるように実装している。その中でも最も早く出現した動物体には、フォーカスを当てている。ユーザはフォーカスが当たっている動物体に対して、あらかじめユーザが登録したラベルに割り当てたキーを入力することで、素早くアノテーションを付与していくことができる。

付与したラベルデータは、ラベル確認ウィンドウで確認することができる(図 3)。確認ウィンドウでは、ラベル付けした(合成前の)動画を、ラベルと共に再生できる。ラベル付けの対象となった動物体が矩形で囲われ、その中央部に、付与したラベルが表示される。ユーザはラベル確認ウィンドウでラベルを確認しながら、ラベルの修正や削除を行うことが可能だ。

4. 議論

4.1 アノテーションのしやすさとコンテキストの維持

セグメントではなく、軌跡データ単位で重畳を管理し、決まった数の軌跡データを常に重ね合わせていくことも可能である。この場合、合成動画上で表示される動物体の数が常に一定となるため、ユーザは、常に自分が処理しているやすい数の動物体を見ながらラベル付けを行っていくことができる。しかし、その分、同時に存在した動物体同士の位置関係というコンテキストが損なわれる。

また、コンテキストという観点では、セグメント単位で重畳した場合でも、重畳するセグメントが多くなってくる

と、背景画像と動物体との位置関係というコンテキストが失われる場合がある。現在の実装では、複数枚のフレームの各ピクセルにおける最頻値を用いて背景を動的に生成し、重畳するセグメントのもっとも古い時刻に合わせて生成した背景を合成している。産卵のターゲットとなる貝は、ほとんど動きが無いため背景として表示されるが、まれにゆっくりと動くことがある。貝が動いてしまった場合、別の時間帯のセグメントにおける動物体の位置と貝との位置関係は崩れてしまう。

ラベル付けをする人は、ラベル付けによって何を確認したいかという観点から、ラベル付けのし易さとコンテキストの維持の双方を考慮し、セグメント単位/軌跡データ単位の重畳の判断、重畳するセグメント数 or 軌跡データ数の決定を行う必要がある。個別の動物体にラベルを付けたいのであれば、軌跡データ単位の重畳でも良いが、動物体間の相互作用に対してラベルを付けたいのであれば、セグメント単位で重畳を行うべきだろう。

4.2 対象となる映像

本稿では魚の撮影を主眼とした水中カメラ映像への適用例を紹介したが、RACRA は魚の映像以外にも様々な生物の行動を撮影した映像に利用できる。たとえば、樹液がでている木の幹をターゲットに撮影して、その樹液部に集まる虫を長期間に渡って撮影した映像のラベル付け(ex. “カブトムシ・オス”，“オオムラサキ・メス”等)も高速に行うことができるだろう。一方、例えば、蜂の巣映像のように、もともと大量に生物が映り続けているような映像の場合、複数のセグメントを重畳しなくてもすでに映像中にラベル付けターゲットが常時多数表示されているため、本ツールの効果は薄いと思われる。

また、本稿では computational ethology のツールとして RACRA を位置づけたが、RACRA 自体は定点で撮影された様々な映像に利用可能である。たとえば、大学校内を定点撮影する監視カメラの映像に対するラベル付けや、交差点を監視する映像に対するラベル付けにも利用できる。現状の画像認識技術では、人や車の抽出はある程度可能であるものの、人の性別や年齢層、車の車種やメーカーなどのラベル付けはまだまだ困難である。このような場面の状況を詳しく調査するためのツールとしても利用できると考えている。

5. Future Work

希少魚生息池で撮影した映像は、日照条件の変化や水の濁り、浮遊藻など、様々な画像処理を困難にする要素が重なり、RACRA が前提とする動物体の軌跡データ抽出に課題を抱えている。現在は、単純な閾値に基づく背景差分等を利用しているが、今後軌跡データの抽出により高度な画像処理を検討する予定である。また、ターゲットによって

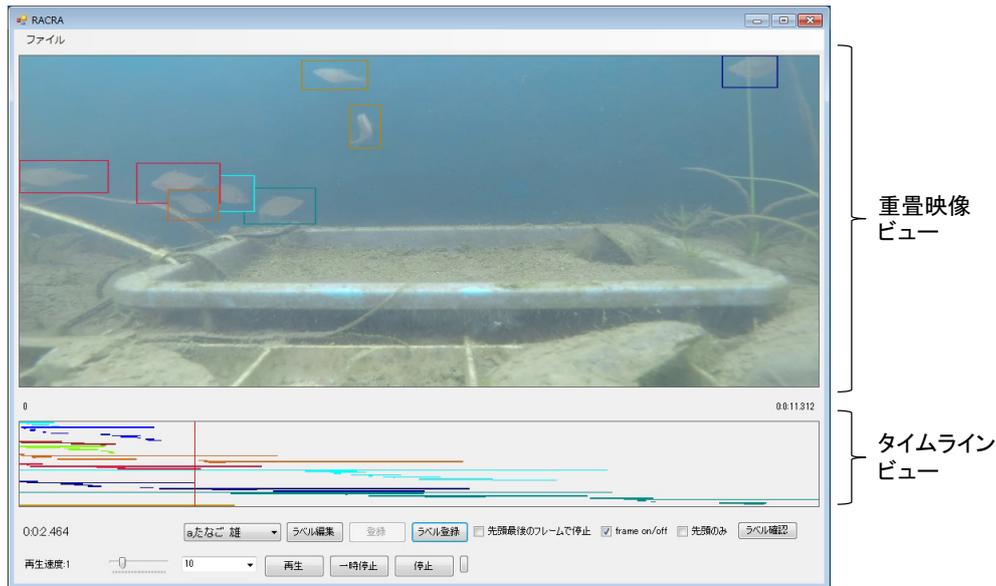


図 2 RACRA のメインウィンドウ
Fig. 2 RACRA: main window



図 3 ラベル確認ウィンドウ
Fig. 3 RACRA: label confirmation

は、ステレオカメラやサーモカメラを利用し、軌跡データの抽出を簡単にする方法も考えられる。

一方、誤検出された軌跡データの修正も RACRA の画面上で簡易に実現できないかについても検討している。たとえば、異なる別々の軌跡と認識されてしまった動物体をマウス操作によって一つの軌跡データとする等のことは実現可能であろう。

6. おわりに

本稿では、定点カメラで撮影された映像に対する高速なラベル付けを可能とするツール、RACRA について述べた。今後、軌跡データ抽出を改良し、適用事例で述べた水中撮影映像に対するラベル付けを進める予定である。

参考文献

[1] Anderson, D.J., Perona, P.: Toward a science of computational ethology, *Neuron*, 84(1), pp. 18-31, 2014.

[2] Bianco, S., Ciocca, G., Napolitano, P., Schettini, R.: An interactive tool for manual, semi-automatic and automatic video annotation, *Computer Vision and Image Understanding*, Vol. 131, pp. 88-99, 2015.

[3] Cockburn, A., Gutwin, C., Alexander, J.: Faster document navigation with space-filling thumbnails. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*, pp. 1-10, 2006.

[4] 福永 津嵩, 岩崎 渉: Computational Ethology: バイオインフォマティクスと動物行動学の融合, 領域融合レビュー, 4, e003, 2015.

[5] Fukunaga, T., Kubota, S., Oda, S., Iwasaki, W.: GroupTracker: Video Tracking System for Multiple Animals under Severe Occlusion. *Computational Biology and Chemistry*, 57, pp. 39-45, 2015.

[6] Kishino, Y., Yanagisawa, Y., Shirai, Y., Mizutani, S., Naya, F., Kitagawa, T.: A Habitat-monitoring System for an Endangered Fish Using a Sensor Network, *UbiComp/ISWC'15 Adjunct*, pp. 133-136, 2015.

[7] Mihalcik, D., Doermann, D.: The design and implementation of viper, Technical Report, University of Maryland, 2003.

[8] 水谷 伸, 白井 良成, 岸野 泰恵, 柳沢 豊, 大津賀 真之, 北川 忠生, 納谷 太: 絶滅危惧の魚の繁殖生態解明に向けたオンライン画像行動解析の予備的検討, マルチメディア, 分散, 協調とモバイル (DICOMO2016) シンポジウム, pp.572-576, 2016.

[9] 坂田 伊織, 岡田 龍也, 杉本 智嗣, 須山 敬之, 柳沢 豊, 岸野 泰恵, 松永 賢一, 北川 忠生: センサネットワークによるニッポンバラタナゴの生息環境のモニタリングシステムの構築, 近畿大学農学部紀要, No.47, pp. 77-85, 2014.

[10] UMATracker, <http://ymnk13.github.io/UMATracker/>.

[11] Vondrick, C., Patterson, D., Ramanan, D.: Efficiently Scaling Up Crowdsourced Video Annotation, *International Journal of Computer Vision (IJCV)*, 101(1), pp. 184-204, 2012.

[12] Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H.: ELAN: a Professional Framework for Multimodality Research, *Fifth International Conference on Language Resources and Evaluation*, 2006.