AI 特許紹介(57)
AI 特許を学ぶ！究める！
～反復ネットワーク特許～

2023 年 10 月 10 日

河野特許事務所
所長弁理士　河野英仁

「AI 特許紹介」シリーズは、注目すべき AI 特許のポイントを紹介します。熾烈な競争となっている第４次産業革命下では AI 技術がキーとなり、この AI 技術・ソリューションを特許として適切に権利化しておくことが重要であることは言うまでもありません。

AI 技術は Google, Microsoft, Amazon を始めとした IT プラットフォーマ、研究機関及び大学から毎週のように新たな手法が提案されており、また AI 技術を活用した新たなソリューションも次々とリリースされています。

本稿では米国先進 IT 企業を中心に、これらの企業から出願された AI 特許に記載された AI テクノロジー・ソリューションのポイントをわかりやすく解説致します。

1.概要
特許出願人　Google
出願日　2020 年 6 月 10 日
公開日　2021 年 12 月 16 日
公開番号 WO2021251959
発明の名称　時間的自己類似性行列を利用したビデオ内のクラスに依存しない繰り返しカウント

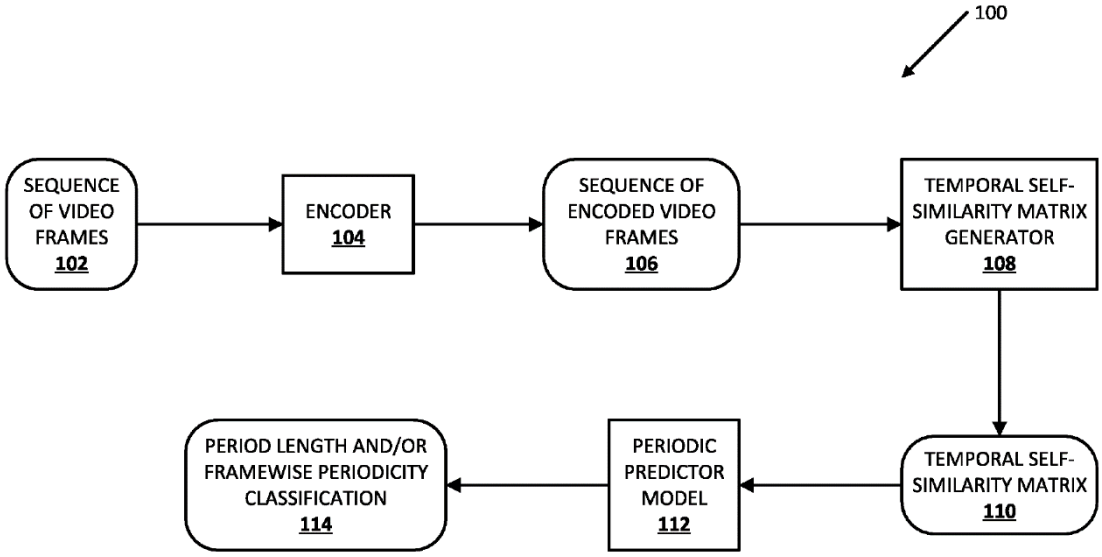
959 特許は、周期的な活動をキャプチャした映像を、反復ネットワークを使用して処理し、期間長および周期性分類を推定する技術に関する。

2.特許内容の説明
カフェにおいて、一人で食事をしている人を想像してみると、食べ物を咀嚼しながらコーヒーに砂糖を入れ、BGM に合わせて足をトントンと叩いている。この人は、少なくとも 3 つの周期的な活動を並行して行っている。

日常生活には、繰り返しの行動やプロセスがよく見られる。959 特許は、ビデオ内でアクションが繰り返される期間を推定するための手法を対象としている。このアプローチの核心は、実際のビデオの目に見えない繰り返しへの一般化を可能にする中間表現のボトルネックとして時間的自己類似性を使用し、期間予測モジュールを制約することにある。

また、このモデルは、さまざまな長さの短いクリップをサンプリングし、それらをさまざまな期間と回数で繰り返すことによって、大規模なラベルのないビデオコレクションから生成された合成データセットを使用してトレーニングを行う。合成データと強力だが制約のあるモデルを組み合わせることで、クラスに依存しない方法で周期を予測できる。

下記図は、反復ネットワークを使用してビデオフレームのシーケンスを処理し、期間長および周期性分類を生成する例 100 を示す。



ビデオフレーム 102 のシーケンスは、エンコーダ 104 を使用して処理され、エンコードされたビデオフレーム 106 のシーケンスを生成する。ビデオフレーム 102 のシーケンスは、周期的な活動（例えば、羽ばたく鳥、人間の心臓の鼓動、一杯のコーヒーの中の砂糖をかき混ぜる人間など）をキャプチャする。

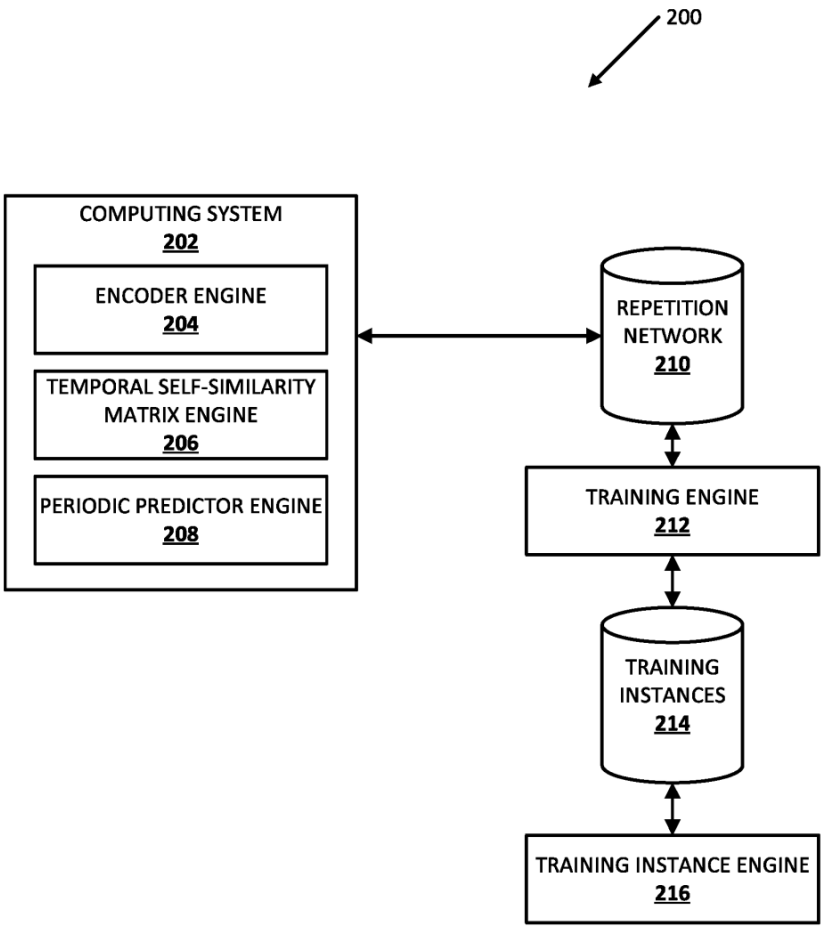
エンコーダ 104 は、ビデオフレームシーケンス 102 の各フレームをエンコードする。例えば、ビデオフレーム 102 のシーケンスは、エンコーダの畳み込みニューラルネットワーク部分を使用して処理され、各ビデオフレームの 2 次元畳み込み特徴を生成する。各ビデオフレームの二次元畳み込み特徴は、エンコーダの三次元畳み込みニューラルネ

ットワーク部分を使用して処理され、各ビデオフレームの時間コンテキスト特徴を生成する。さらに、各ビデオフレームの時間コンテキスト特徴は、エンコーダの最大プーリング部分を使用して処理され、エンコードされたビデオフレームを生成する。

エンコードされたビデオフレームのシーケンス 106 は、時間的自己類似度行列 110 を生成するために、時間的自己類似度行列生成器 108 を使用して処理される。時間的自己類似性行列 110 は、エンコードされたビデオフレームのシーケンス 106 における、エンコードされたビデオフレームのペア間のペアごとの類似性である。例えば、自己類似度行列 110 は、時間的自己類似度行列生成器 108 を使用して、エンコードビデオフレーム 106 のシーケンスにおけるエンコードビデオフレームのすべての対の間の負の二乗ユークリッド距離を生成することによって生成する。エンコードされたビデオフレームのすべてのペア間の負の二乗ユークリッド距離を生成した後に、行単位のソフトマックス演算を実行できる。

時間的自己類似性行列 110 は、ビデオフレーム 102 のシーケンス内でキャプチャされた周期的活動の周期長、および、ビデオフレームシーケンス 102 内の各ビデオフレームが周期的アクティビティをキャプチャするかどうかを示すフレームごとの周期性分類 114 を生成するために、周期予測子モデル 112 を使用して処理される。

下記図は、環境 200 のブロック図を示す。



環境 200 は、エンコーダエンジン 204、時間的自己類似性行列エンジン 206、期間予測エンジン 208 及びコンピューティングシステム 202 を含む。コンピューティングシステム 202 は、反復ネットワーク 210、トレーニングエンジン 212、トレーニングインスタンス 214、及びトレーニングインスタンスエンジン 216 と関連付けられる。

上記図に示されるように、トレーニングインスタンスエンジン 216 は、トレーニングインスタンス 214 を生成するために使用される。トレーニングインスタンスエンジン 216 は、ラベルなしビデオを使用して、合成トレーニングインスタンス 214 を生成するために使用される。

トレーニングインスタンスエンジン 216 は、下記図で説明される合成ビデオに基づいてトレーニングインスタンス 214 を生成することができる。トレーニングエンジン 212 は、反復ネットワーク 210 をトレーニングするために使用される。トレーニングエンジン 212 は、トレーニングインスタンス 214 のうちの 1 つまたは複数を処理して、トレーニング損失を生成することができ、トレーニング損失は、反復ネットワーク 210 の 1 つまたは複数の部分を、例えば誤差逆伝播法により更新するために使用される。

例えば、トレーニングインスタンス 214 は、トレーニング周期的アクティビティおよびグランドトゥルース周期データをキャプチャするトレーニングビデオを含むことができ、グランドトゥルース周期データは、トレーニング周期的アクティビティの期間長、トレーニングビデオのフレームごとの周期性表示、およびトレーニングビデオ内の周期的なアクティビティの繰り返しの数を含む。

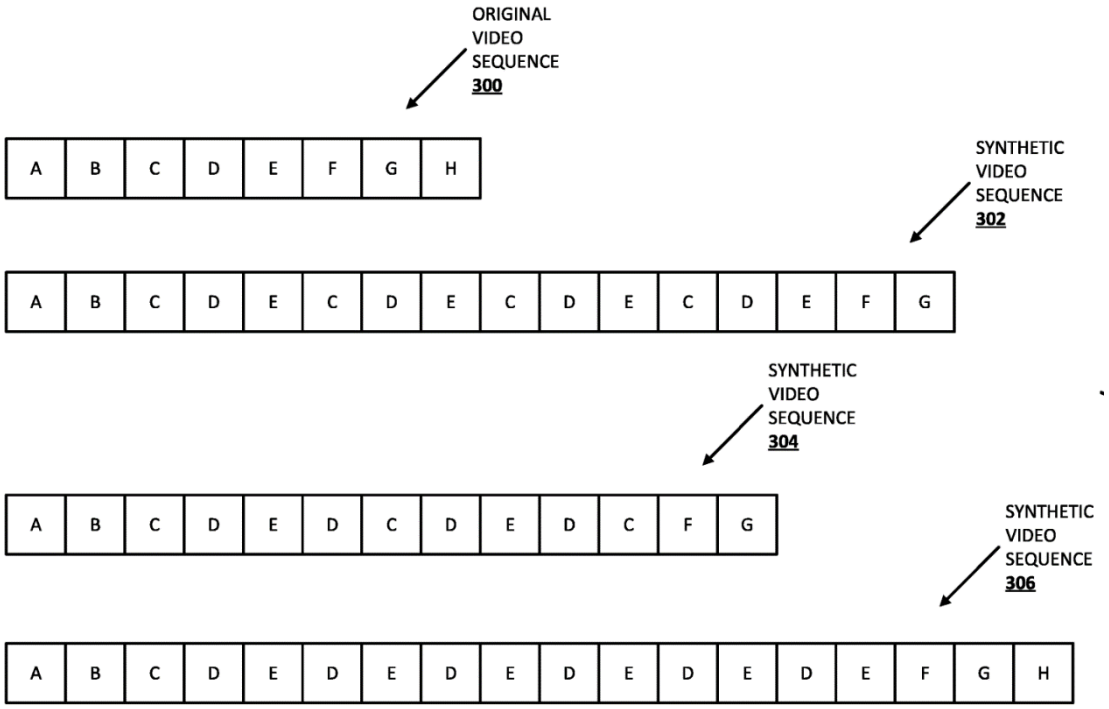
エンコーダエンジン 204 は、反復ネットワーク 210 のエンコーダ部分を使用してビデオフレームのシーケンスを処理し、ビデオフレームのエンコードされたシーケンスを生成するために使用することができる。エンコーダエンジン 204 を使用して、エンコーダ 104 を使用してビデオフレーム 102 のシーケンスを処理し、エンコードされたビデオフレーム 106 のシーケンスを生成することができる。

時間的自己類似性行列エンジン 206 は、エンコーダエンジン 204 を使用して生成されたエンコードビデオフレームのシーケンスなど、エンコードされたビデオフレームのシーケンスを処理して、時間的自己類似性行列を生成するために使用される。時間的自己類似性行列は、ビデオフレームの符号化シーケンスにおけるエンコードビデオフレームの各ペアのペアごとの類似性である。

時間的自己類似性行列エンジン 206 は、時間的自己類似性行列生成器 108 を使用して、エンコードされたビデオフレーム 106 のシーケンスを処理して、時間的自己類似性行列 110 を生成する。

周期予測エンジン 208 は、反復ネットワーク 210 の周期予測モデル部分を使用して、ビデオフレームのシーケンス（すなわち、エンコーダエンジン 204 を使用して処理されるビデオフレームのシーケンス）内にキャプチャされた周期的活動の期間長、及び、ビデオフレームのシーケンスのフレームごとの周期性分類を生成するために使用される。周期予測子エンジン 208 は、周期予測子モデル 112 を使用して、ペアごとの長さおよびフレームごとの周期性分類 114 を生成するために、時間的自己類似性行列 110 を処理する。

下記図は、元のビデオシーケンス 300 に基づいて生成された合成ビデオ 302、304、および 306 の例を示す。



図の例では、元のビデオシーケンス 300 は、'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H'の８つのビデオフレームのシーケンスを含む。合成ビデオシーケンス 302 は、オリジナルビデオシーケンス 300 に基づいて、'C', 'D', 'E'のオリジナルビデオシーケンスの３フレーム部分を選択することによって生成された例示的な合成ビデオシーケンスである。

選択された'C', 'D', 'E'の 3 つのフレーム部分が 4 回繰り返され、3 回繰り返されるアクティビティを表す。さらに、'C', 'D', 'E'の選択された３フレーム部分の直前にある、'A'、'B'の元のビデオの２フレーム部分が、合成ビデオシーケンス 302 の先頭に付加される。同様に、'C, 'D', 'E'の選択された３フレーム部分の直後の元のビデオ'F', 'G'の２フレーム部分が、合成ビデオシーケンス 302 の最後に追加される。

合成ビデオシーケンス 304 は、'C', 'D', 'E'の３つのフレームシーケンスを選択することによってオリジナルビデオシーケンス 300 に基づいて生成される別のサンプル合成ビデオシーケンスである。さらに、選択されたシーケンスが'D', 'C'と反転される。選択された３つのフレームシーケンスおよび'C', 'D', 'E', D', 'C'の反転シーケンスは、合成ビデオシーケンス 304 内で２回繰り返され、アクションが繰り返されている間に反転されるジャンピングジャックなどの周期的なアクティビティを表す。

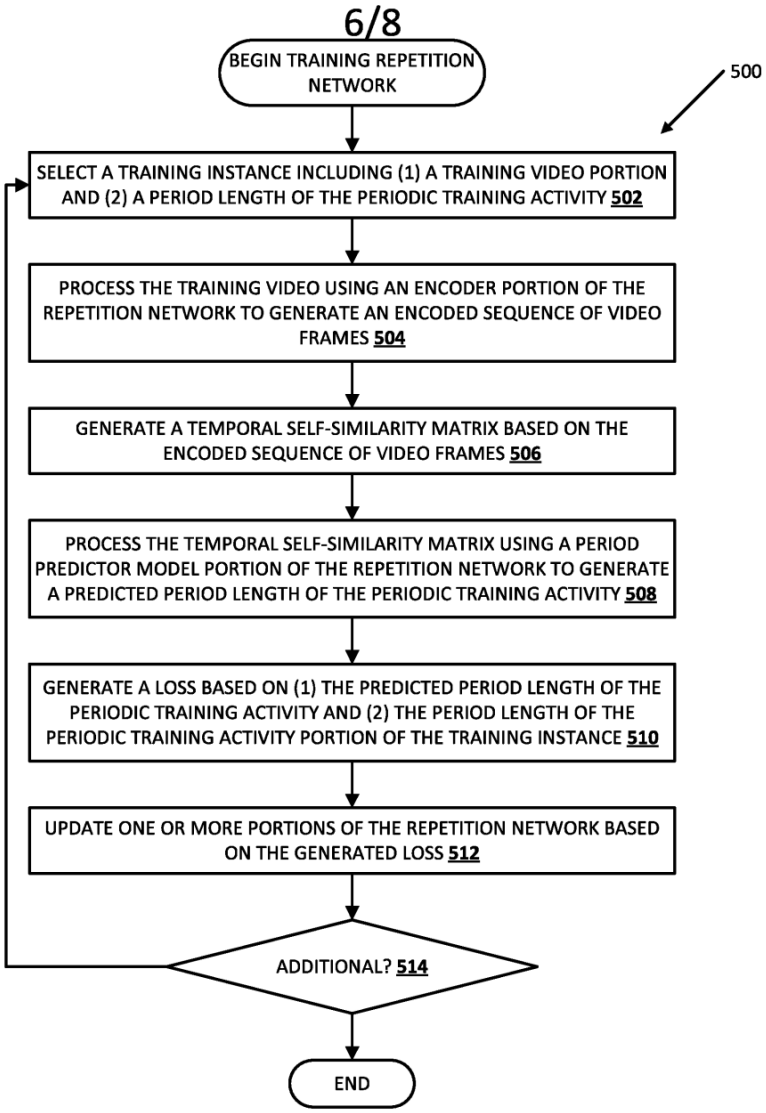
'C', 'D', 'E'の選択された３フレーム部分の直前にある 'A', 'B'の２フレーム部分は、合

成ビデオシーケンス 304 の先頭に付加される。同様に、'C', 'D', 'E'の選択された 3 フレーム部分の直後の元のビデオの 2 フレーム部分'F', 'G'が、合成ビデオシーケンス 304 の最後に追加される。

合成ビデオシーケンス 304 は、オリジナルビデオシーケンス 300 に基づいて生成されたさらなるサンプル合成ビデオシーケンスである。合成ビデオシーケンス 304 は、'C', 'D', 'E', 'F'の 4 つのフレームシーケンスを選択することによって生成される。さらに、選択された 4 つのフレームシーケンスが 'E', 'D', 'C'のように反転される。

選択された 4 フレームシーケンスおよび反転シーケンスは、合成ビデオシーケンス 306 内で 4 回繰り返され、シーケンスの反転を含む周期的な活動（例えば、ジャンピングジャック）を表す。選択された 4 つのフレームシーケンスの直前のビデオフレーム'C'は、合成ビデオシーケンス 306 の先頭に付加される。さらに、選択された 4 つのフレームシーケンスの直後のビデオフレーム'G'が、合成ビデオシーケンス 306 の最後に追加される。

下記図は、反復ネットワークを訓練するプロセス 500 を示すフローチャートである。



ブロック 502 で、システムは、周期的なトレーニング活動をキャプチャするトレーニングビデオおよびグランドトゥルースの周期的な出力を含むトレーニングインスタンスを選択する。グランドトゥルースの周期的出力には、（1）周期的トレーニングアクティビティの周期長、（2）トレーニングビデオ内の周期的トレーニングアクティビティの繰り返し数、および（3）トレーニングビデオのフレームごとの周期性の表示が含まれる。

ブロック 504 で、システムは、反復ネットワークのエンコーダ部分を使用してトレーニングビデオを処理し、ビデオフレームのエンコードシーケンスを生成する。ブロック 506 で、システムは、ビデオフレームのエンコードシーケンスに基づいて時間的自己類似性行列を生成する。

ブロック 508 で、システムは、反復ネットワークの周期予測子モデル部分を使用して時間的自己類似性行列を処理し、予測周期出力を生成する。ブロック 510 で、システムは、（1）予測された周期出力、および（2）トレーニングインスタンスのグランドトゥルース周期出力部分に基づいて損失を生成する。ブロック 512 で、システムは、発生した損失に基づいて、反復ネットワークの 1 つまたは複数の部分を（例えば逆伝播を通じて）更新する。以上の処理によりトレーニングされた反復ネットワークを用いて、入力されたビデオの周期長及び繰り返し数等を予測する。

3.クレーム

959 特許のクレーム 1 は以下の通りである。

1.　1 つ以上のプロセッサによって実装される方法において、

エンコードされたビデオフレームのシーケンスを生成するために、反復ネットワークのエンコーダ部分を使用して、周期的なアクティビティをキャプチャしたビデオフレームのシーケンスを処理し、

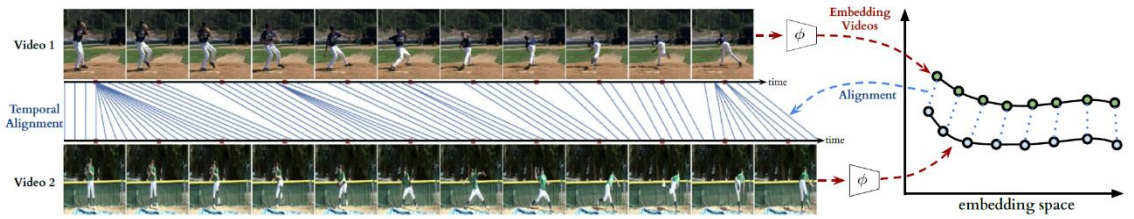
エンコードされたビデオフレームのシーケンスに基づいて、エンコードされたビデオフレームのシーケンス内のエンコードされたビデオフレーム間のペアごとの類似性を示す時間的自己類似性行列を生成し、

(a)ビデオフレームのシーケンスにおける周期的アクティビティの周期長、及び/または(b)ビデオフレームのシーケンスのフレームごとの周期性分類を生成するために、反復ネットワークの周期予測子モデル部分を使用して時間的自己類似性行列を処理する。

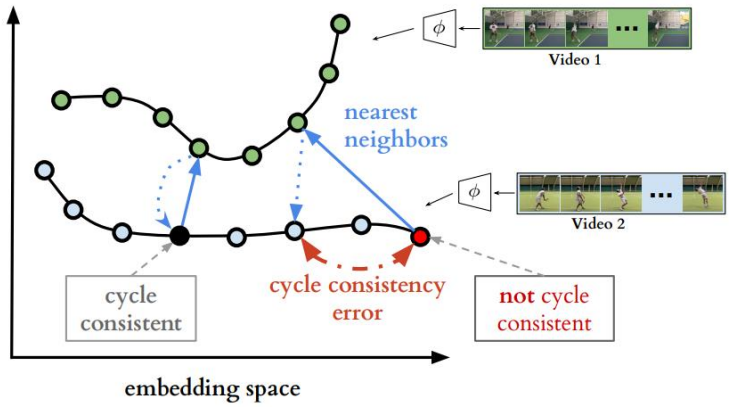
4. 本特許に関連する論文

本特許に関する論文“Temporal Cycle-Consistency Learning”[1]が、Google Brain 及び DeepMind の Debidatta Dwibedi 氏らにより公表されている。本論文と 959 特許とはビデオの周期性を推定するという点で共通するが、本論文は 2 つの異なるビデオの対応関係を推定する時間的サイクル一貫性（TCC：Temporal Cycle Consistency）学習と呼ばれる自己教師あり表現学習手法を紹介している。

[1] Debidatta Dwibedi et al. “Temporal Cycle-Consistency Learning” arXiv:1904.07846v1 [cs.CV] 16 Apr 2019



この方法では、微分可能なサイクル一貫性損失である時間的サイクル一貫性（TCC）を使用してネットワークをトレーニングする。TCC は、複数のビデオの時間にわたる対応関係を見つけるために使用される。結果として得られるフレームごとの埋め込みは、学習された埋め込み空間で最近傍を使用してフレームを一致させるだけで、ビデオの位置合わせを行うことができる。



上記図は、埋め込み空間の例でエンコードされた 2 つのビデオシーケンスの例を示す。マッチングに最近傍を使用する場合、1 つの点（黒で表示）はそれ自体に循環して戻るが、別の点（赤で表示）は循環して戻らない。最大数のポイントが循環して自身に戻ることができる埋め込み空間を学習する。これは、シーケンスの各ペアの各ポイントのサイクル一貫性エラー（赤い点線で表示）を最小限に抑えることで実現される。なお、Youtube に TCC の動画が紹介されている。

https://www.youtube.com/watch?v=iWjjeMQmt8E

以上

著者紹介

河野英仁

河野特許事務所、所長弁理士。立命館大学情報システム学博士前期課程修了、米国フランクリンピアースローセンター知的財産権法修士修了、中国清華大学法学院知的財産夏

季セミナー修了、MIT(マサチューセッツ工科大学)コンピュータ科学・AI 研究所 AI コース修了。

AI 特許コンサルティング、医療 AI 特許コンサルティングの他、米国・中国特許の権利化・侵害訴訟を専門としている。著書に「世界のソフトウエア特許(共著)」、「FinTech 特許入門」、「AI/IoT 特許入門 3」、「ブロックチェーン 3.0(共著)」がある。