

論文の内容の要旨

Visual Cue Selection, Substitution and Integration for Robust Object Tracking

ロバストな物体追跡のための視覚的特徴の選択と統合に関する研究

パスカル アジタ アセラ

目的

本研究の主な目的を以下に示す。

- 複数の視覚特徴 (Visual Cue) を用いた実世界シーンのロバストな物体トラッキング手法の研究とその開発
- ロバストな物体トラッキングのために簡単で効率的な手法として視覚特徴の選択と統合に関する 研究

多くのコンピュータ・ビジョンシステムにおいて、物体トラッキングは重要な役割を担っており、その結果が他の外部システムの入力として扱われることが多い。様々な動的な変化が伴う実世界におけるトラッキングは多くの課題が残されている。制御された環境で、例えば照明条件が一定で、動く物体が一個に限定されている場合の物体トラッキングは比較的簡単である。しかし、文献によれば、一つの視覚特徴を用いた場合、比較的低い複雑さの環境においてもトラッキングの失敗が生じる。この問題を解決するため複数の視覚特徴 (多くの場合並列的に) を用いることが提案されており、それらによって計算された物体位置の推定はどの視覚特徴もそれ単独で用いた場合よりも精度良く求められる。しかし、ここで主に問題となるのは適切な統合法を決定することである。文献では統計的手法と投票に基づく方法の二つの基本統合法が示されている。さらに、視覚特徴統合にファジー理論を用いる手法も報告されている。

本分野の研究は活発に行われる一方で、ターゲットについて事前の知識を持たない一般的な物体のトラッキング・アルゴリズムについての研究はあまり例をみない。殆どの物体トラッキング文献は既知の物体を扱っている。一般的な物体と実環境の下という二つの条件を満たす物体トラッキングでは、どのような視覚特徴を用いるか、どのような枠組みの中でそれらを用いるかが問題となる。本論文では上記の問題を扱い、複数の視覚特徴に基づき、一般的な物体のトラッキングにおける新しい枠組みを提案する。本論文では視覚特徴選択、自動代用そして統合に基づいている。

視覚特徴選択

物体について事前知識がない場合の視覚特徴選択に関する従来の研究は、我々の知る限り

では存在しない。本手法で視覚特徴選択は二つの段階で行われる。まず、最初に一般的な未知の物体に関して、どのような視覚特徴を使うかという問題がある。ここでは、初期フレームが分析され、適切な特徴の順位付けが行われる。例えば、カラー情報のように事前に分かっている情報を利用できる。この段階では、ターゲット物体の統計的な性質に基づくパラメータが各視覚特徴について計算され、(参照画像として与えられる、あるいは一つの動く物体のみの場合は自動的に抽出される)これらのパラメータはトラッキングに利用できるかどうかの判断が行われる。テクスチャー / 濃淡情報の場合は、その白黒ヒストグラムの差分の分散が計算される。その分布の間十分の差があれば、テクスチャー / 濃淡情報は一つの候補となる。Tomasaiによって提案されている *textureteness* 指標を用いてターゲットのトラッキングに利用できる十分なテクスチャーを持っているかを調べる。カラー情報に関しては画像のカラー・ブロック (color blob) も利用する。ある範囲の速度を持っている画素と対象シーンに動く物体がいくつあるかということを用いて物体トラッキングにオプティカル・フローの適用も検証する。

現在用いている視覚特徴によるトラッキングが失敗する場合にそれを代用する特徴を多くの中から選択するのは更に困難な問題である。ここで、データ連想手法と各視覚特徴(各々の特徴により独立にターゲット位置を予測する)に伴う共分散行列、それらから計算された重みを用いてどの特徴が最も良いかを決定する。

視覚特徴選択と統合

本論文の提案手法では、主な視覚特徴という概念を提案し、単独の特徴で不確実性が高くなる(トラッキングが失敗する)までトラッキングを行う。視覚特徴の統合はトラッキングが不確定の時間から次の主な視覚特徴を見つけるまでのみ行われる。それから再び一つの視覚特徴でトラッキングを行う。現在の主な視覚特徴から他の特徴への代用には他の視覚特徴の情報を用いる。この動作は自動であり視覚特徴代用と呼ぶ。本手法の一つ重要な性質はある瞬間においてはひとつの特徴のみでトラッキングが行われるということである。

複数の特徴は一つの特徴のみで困難な複雑な環境におけるトラッキングの不確実性を除去するために用いる。

本手法は、エラーリカバリも含んでおり、提案の枠組みのなかで、複雑な環境において複数の視覚特徴を扱い、物体のロバストなトラッキングを行うシステムが実現できる。各視覚特徴のみを用いる独立のトラッキング装置にそれらをコーディネートするひとつのコントローラの追加により上記が可能となる。

また、提案枠組みはあらゆる視覚特徴が扱えるように拡張できる(ここで扱っていない、

物体への距離 そして 物体の形状なども含む)。本概念の有効性を示すために本論文では、テクスチャー / 濃淡情報、色情報、動きと エッジ情報を用いている。もう一つの重要な要素は比較的複雑な実世界動画像系列を扱い、照明条件の変動、複雑な背景、オクルージョンと近い色の物体の存在などのノイズ的な要因の中でその有効性を証明したことである。このような動画像系列を用いることは今後のトラッキング手法のロバストさの検証に不可欠だと考える。

本手法のアプリケーションとして複数カメラを用いた顔トラッキングシステムを構築した。本システムにおいては、一つのコンピュータに複数のカメラを接続し、ある瞬間においては一つのカメラのみが動作している。ターゲットは一つのカメラの視界から離れると顔の動きに対応して次のカメラが動作する仕組みになっている。

複数カメラを用いた実時間顔トラッキングシステム

上記のトラッキング枠組みに基づく顔実時間トラッキングシステムを複数のカメラを用いて実現した。本システムでは、広い領域における顔の動きに対応するためにカメラを機械的に制御する必要がないためシステムの複雑さが削減できる。隣接するカメラの間で視界をある程度オーバーラップさせることで一つのカメラから別のカメラに動作が移る場合、滑らかな顔トラッキングを可能としている。結果としてビデオ・レートの顔トラッキングが実現できた。

システムの第一段階としては、顔の位置のみが計算される。カメラの位置を決定するにあたり、隣接した二つのカメラの間に十分なオーバーラップ領域があるかが基準となる。カメラの台数に関してはカバーすべき領域によって決まり、顔からカメラへの距離はどのアプリケーションによって決定される。ここで、顔の動きはカメラがある方向へのものに限定される。

顔は一つのカメラの視界から別のカメラへ移る領域に存在し、そのカメラの方向へ動くとき、自動的に作動する。そして第二のカメラが第一カメラのトラッキング情報に基づいて大まかな領域の中で顔トラッキングを始める。それによってカメラ並びの位置による影響をなくす。

顔トラッキングアルゴリズム自身がカラー、動きと エッジ情報を視覚特徴として用いている。