

平成 19 年度 情報工学コース卒業研究報告要旨

大西 研究室	氏 名	刈 込 学
卒業研究題目	動画からの特定動作検索	

背景と目的

本研究では、入力動画の中から、特定の動作を検出するシステムを作成することを目的とする。このようなシステムは、例えば、スーパーマーケット等に設置された防犯カメラから、万引き行為等の特定の動作を自動的に検出することに応用できる。他にも、看護ロボットの画像認識システム等にも応用できると思われる。

入力は、人間の動作が含まれる、シルエット画像列となる（入力動画）。この入力動画の中から、検出したい動作（検出動作）に近い部分を見つけることになる。検出動作とは、入力動画に含まれているかどうか検出したい動作の、エッジ画像列である。検出動作は、右手を挙げる、しゃがむ、などの、短く単純な動きに限定する。ダンスなど、多様な動きの含まれる動作については検出の対象外とする。

特定動作検出の処理

画像の類似度を測る尺度として、Chamfer 距離を用いる。Chamfer 距離による類似画像検索は、計算コストが小さくてすむ、という利点がある。しかし、Chamfer 距離による類似画像検索では、オクルージョンが発生している場合など、特定の画像に対して検出率が落ちる問題がある。この問題に対する解決策として、キーフレームを用いる。キーフレームとは、検出したい動作の中で、極力オクルージョンが発生しておらず、その動作特有の姿勢の画像である。このような画像を各動作に一枚設定しておくことで、検出率を上げることができる。

検出処理の流れは、以下のようになる。まず、キーフレームから入力動画の各フレームへの Chamfer 距離を計算し、リスト化する。最小の Chamfer 距離のものが、入力動画でもっともキーフレームに近いフレームと判断する。その後、検出動作の開始部分と、終了部分の姿勢に近い画像を検出する。キーフレームの近くに先頭画像、最後の画像は存在するはずなので、検出範囲を制限する。制限した範囲内で、先頭画像、及び最後の画像の Chamfer 距離リストを作成する。それぞれ、最小 Chamfer 距離のフレームと、最小 Chamfer 距離から差が 0.1 以下の画像が開始部分候補、終了部分候補となる。開始部分と終了部分に近い画像間が、検出動作をしている可能性が高い。これが入力動画における、検出動作候補部分となる。

次に、検出動作の重ね合わせ画像と、検出動作候補部分の重ね合わせ画像を作成する。同じ動作をしているなら、重ね合わせ画像同士も近いものになるはずである。検出動作の重ね合わせ画像から、検出動作候補部分の重ね合わせ画像への Chamfer 距離を求める。各検出動作候補部分に対して、この操作を行う。最小 Chamfer 距離のものが、検出動作に最も近い部分と判断し、それを出力する。

実験と結果

検出動作として 10 種類のエッジ画像列と、それら 10 種類の動作を含む、516 フレームの入力動画を用意した。この条件下で、入力動画内から、検出動作と類似する動作を検出した。検出は、キーフレームを用いる処理と用いない処理と、2 種類の処理で行った。キーフレームを用いる処理は 10 種類すべて検出できた。一方、キーフレームを用いない処理では、検出できたのは 9 種類であり、1 種類誤検出した。実験結果から、以下の式であらわされる、重なり率を求めた。

$$(\text{重なり率}) = \frac{\text{入力動画から検出された候補のうち、検出動作と重なるフレーム数}}{\text{検出動作のフレーム数}} \quad (1)$$

重なり率の平均は、キーフレームを用いた処理では 83.6%、キーフレームを用いない処理では 65.4%であった。また、計算速度はキーフレームを用いることで 3.5 倍向上した。