

## 博士学位論文概要

論文題目	ピッキングと鍵開けにみる類似行動を 機械学習により識別する研究
執筆者	白石 将 貴
指導教員	宇田 隆 哉 講師
<p>本論文は、家人による解錠操作と、ピッキングによる解錠操作を区別することを目的とするものである。住宅のセキュリティを考える際、出入り口以外は一般的に人が出入りしないため、簡易なセンサによって不正な侵入を検知することが可能であるが、出入り口は家人も出入りするため、単純なセンサによる検知はできない。監視カメラの設置は一般的であるが、侵入があったことを後日知ることはできても、侵入時にその侵入を検知するには有人監視を常時行うしかない。そこで、本論文では、Kinect を用いて不正な侵入を自動的に検知するシステムの提案を行う。なお、Kinect により骨格座標を取得している。Kinect を用いた動作識別の研究は既に行われており、不正な侵入のなかでも、通常の入出とは異なる、ドアを破壊するような行為の検知は既存技術でも対応可能である。そこで、本論文では家人による解錠操作と、ピッキングによる不正な解錠操作を区別する点に重点を置いた。本研究では機械学習を用いている。大量にデータを集めて学習しさえすれば、従来の単純な機械学習でも家人による解錠操作とピッキングを区別できる可能性はあるが、学習にとっても時間が掛かりコストが増加する。そこで、提案手法では、少人数の被験者のデータから、体格による分布が均一になるようにサンプルを増加させ、少ないデータで効果的に学習が行えるように工夫した。Kinect を使用することにより、撮影された画像を保存する必要がないため、被撮影者のプライバシーも保護される。実験の結果、ピッキングを平均 76%、被験者によっては 94%の精度で区別ができた。なお、右腕の指先から肩までの骨格座標を取得した場合には、平均 73%、被験者によっては 97%の精度で区別ができ、ピッキングおよび鍵開け動作を行っている右腕に特徴があらわれていた。</p>	

## Abstract

Title	Research for Classification of Similar Actions by Machine Learning with Examples of Picking and Key-Unlocking
Author	M a s a k i S h i r a i s h i
Supervisor	Senior Assistant Professor R y u y a U d a

The objective of this paper is to distinguish opening a door by the key from opening it by picking. In terms of security of a house, windows can be monitored by small sensors since no one goes through windows in usual. On the other hand, doors cannot be monitored by the sensors since not only thieves but also residents go through the doors. CCTV (closed-circuit television) camera is one of the solutions. However, invasion can be detected not during but after the invasion, or the movie by the camera must be always watched by a person. Therefore, in this paper, I propose a system which automatically detects invasion by Kinect. As well, action coordinates are acquired by Kinect. Of course, there are some researches in which Kinect is used to distinguish human actions, and big actions such as breaking doors can be detected with methods in the researches. In contrast, we put the stress on to distinguish a small difference such as difference between opening a door by the key and opening it by picking. We used machine learning in this paper. If big data of opening a door can be collected, the small difference would be able to be distinguished by usual way of machine learning. However, it takes too much time for learning and the cost increases. Therefore, we made an idea which can learn effectively from small samples of few examinees by increasing samples mathematically. Using Kinect also protects the privacy of people in the movie since only coordinates of joints of people are stored. The average of F-measure of detecting picking was 76% and 94% by an examinee in the results of our experiments. By the way, the average of F-measure of right arm, of detecting picking was 73% and 97% by an examinee in the results of our experiments. And In addition, the right arm performing the distinguish opening a door by the key and opening it by picking was characterized.

## 博士学位論文要旨

論文題目	ピッキングと鍵開けにみる類似行動を 機械学習により識別する研究
執筆者	白石 将 貴
指導教員	宇田 隆 哉 講師
<p>本論文は、家人による解錠操作と、ピッキングによる解錠操作を区別することを目的とするものである。</p> <p>住宅のセキュリティを考える際、出入り口以外は一一般的に人が出入りしないため、簡易なセンサによって不正な侵入を検知することが可能であるが、出入り口は家人も出入りするため、単純なセンサによる検知はできない。また、その他にカードキーを使用する方法もあるが、バッテリーが切れた場合や通電しなくなった場合には、電子ロックが解除されるように設計されており信頼性が低い。したがって、ドアを開錠する最善の方法は、一般的な物理キーを使用することである。監視カメラの設置は一般的であるが、侵入があったことを後日知ることができても、侵入時にその侵入を検知するには有人監視を常時行うしかない。</p> <p>そこで、本論文では、Kinect を用いて不正な侵入を自動的に検知するシステムの提案を行う。Kinect を用いた動作識別の研究は既に行われており、不正な侵入のなかでも、通常の入出とは異なる、ドアを破壊するような行為の検知は既存技術でも対応可能である。そこで、本論文では家人による解錠操作と、ピッキングによる不正な解錠操作という、動作の違いが少ないものを区別する点に重点を置いた。</p> <p>本研究により、類似行動を区別することで、有人監視による警備員の負荷が軽減されると考えられる。具体的には、一戸建て住宅における年間のピッキング侵入件数は11件、年間移動回数より日本国内における鍵開け回数は <math>1.25 \times 10^{11}</math> 回となる。よって、ピッキングと鍵開けの比率は、<math>11 : 1.3 \times 10^{11}</math> となり、正しく鍵開けを検出できるとすると、警備員の負荷が大幅に軽減されることがいえる。</p>	

本研究における被験者の骨格座標を取得する際、Kinect の高さは 185cm とし、胸の高さに近づくようにした。Kinect と被験者の距離は 300cm とし、被験者の全身が中央に収まるようにした。また、Kinect の設置位置は被験者の真横と右斜め後ろで扉に対して 30 度とした。これは、正面もしくは真後ろからが望ましいが、正面にはドアがあり、真後ろだと手元がみえなくなるためである。次に、被験者の立ち方によりばらつきが出ないように被験者 10 名の立ち位置を測定し、その平均とした。かかとの開き幅を 14cm、ドアからかかとの距離を 56cm とした。鍵開けとピッキングの動作もばらつきが出ないように動作を統一した。ピッキングの場合には手で棒状のものを持ち鍵穴付近を上下に動かし、鍵開けの場合には右手で鍵穴に鍵を入れて開け閉めを行う。いずれの動作の場合にも左手はドアノブに手を添えておくこととした。この動作をサンプル取得中連続して行うこととした。なお、連続してデータを 25 サンプル取得する。1 サンプルは 40 フレームから構成されており、2 秒程で取得している。また、1 フレームは各骨格座標の X 座標、Y 座標、Z 座標の値から成っており、立体的に行動を捉えることが可能となっている。

本研究では CNN と SVM を用いている。CNN のネットワークは 11 層とし、畳み込みを 2 回行っている。そして二値分類を行っている。一方、SVM は 3 層とし、分類器には線形 SVC を用いて二値分類を行っている。本論文での精度評価は、Accuracy の値ではなく、F 値としている。なお、CNN による学習の際に必要なバッチサイズやエポック数を事前実験として決定する。その際に、バッチサイズは 32 から倍々に増加させ、エポック数は 50 から 50 ずつ増加させ、F 値が 90% 超かつ標準偏差が 5% 未満のものを使用するものとする。これにより、10 交差検証により精度評価を行い、5 回繰り返し平均精度をとることとする。事前実験での精度評価では、全被験者データが訓練用サンプルとテスト用サンプルの両方に含まれている。一方、本実験の精度評価ではテスト用サンプルに 1 名の被験者データを入れ、残りの被験者データを訓練用サンプルとし、被験者を入れ替えながら 10 回ずつ行っている。

大量にデータを集めて学習しさえすれば、従来の単純な機械学習でも家人による解錠操作とピッキングを区別できる可能性はあるが、学習にとっても時間が掛かりコストも増加する。そこで、提案手法では、少人数の被験者のデータから、体格による分布が均一になるようにサンプルを増加させ、少ないデータで効果的に学習が行えるように工夫した。具体的には、Kinect から取得した骨格座標サンプルを用いて被験者の身長や腕の長さなどの個人特徴量をなくすため、同じ身長となるようにデータ整形を行った。その後、仮想的に体格の異なる被験者のサンプル生成や、線形補間により異なる行動周期のサンプル生成を行った。また、動いている骨格座標が分類に意味があるのか、動いている骨格座標が隠れた場合に分類可能であるのかを分析するために骨格座標の一部を0に置き換え、行動識別を行った。

まず、全骨格座標取得における被験者10名での評価を行った。CNNでは、被験者全体の平均F値は56%と分類は行えていないが、40%台の被験者から80%近くまでいる。しかし、ほとんどの被験者の分類精度は低い。SVMでは、被験者の平均F値は54%と低く、20%台の被験者から70%台の被験者までいる。この原因として、訓練用サンプルの被験者数が少なく、テスト用サンプルの被験者と一致するサンプルが存在しなかったことがいえる。

そこで、次に、被験者10名のデータを用いて、仮想的に体格の異なる被験者を作り出した。この方法により仮想サンプルを増加させ、精度評価を行った。CNNでは、被験者全体の平均F値は59%と分類は行えていない。一番低いF値は41%となっているが、被験者によっては82%、91%と分類できている。SVMでは、被験者全体の平均F値は54%と分類は行えていない。一番低いF値は22%となっているが、被験者によっては80%、81%と分類できている。これは、55名程度の被験者を集めて実験を行えば、一部の高精度を示した被験者に関して精度を向上させることができるが、工夫することで10名の被験者でも精度を向上させることがいえる。そこで、鍵開けとピッキングの行動周期に着目し、類似周期のサンプルを訓練させることで、上手く分類できるのではないかと考えた。

そこで、線形補間により鍵開けとピッキングの行動周期が異なるサンプルを作成し、精度評価を行った。CNNでは、被験者全体の平均F値は57%と分類は行えていない。一番低いF値は35%となっているが、被験者によっては83%、94%と分類できている。SVMでは、被験者全体の平均F値は55%と分類は行えていない。一番低いF値は23%となっているが、被験者によっては81%、85%と分類できている。

次に、被験者数を10名から20名に増加させ、精度評価を行った。CNNでは、被験者全体の平均F値は52%と分類は行えていない。SVMでは、被験者全体の平均F値は76%となっている。また70%台が4名、80%台が2名、90%が2名となっており、被験者数を増加させれば精度が高くなることがわかる。ここで、CNNの精度評価がうまくいかなかったのは、被験者が行動している位置がずれていると考えた。

そこで、被験者の骨格座標の値が近くなるように座標をスライドさせ、精度評価を行った。CNNでは、被験者全体の平均F値は51%と分類は行えていない。SVMでは、被験者全体の平均F値は73%となっている。また80%台が2名、90%が3名となっている。よって、骨格座標をスライドしてもCNNには影響がみられなかった。

次に、右腕の指先から肩までの骨格座標を取得した場合に右肩に特徴があらわれているか評価を行った。CNNでは、被験者全体の平均F値は48%と分類は行えていない。SVMでは、被験者全体の平均F値は73%となっている。また70%台が1名、80%台が4名、90%が1名となっている。よって、右腕の指先から肩までの骨格座標に動作の特徴があらわれていることがわかった。

次に、右腕の指先から肩まで以外の骨格座標を取得した場合の評価を行った。CNNでは、被験者全体の平均F値は48%と分類は行えていない。SVMでは、被験者全体の平均F値は45%と分類は行えていない。

次に、右腕の指先から肘まで以外の骨格座標を取得した場合の評価を行った。CNNでは、被験者全体の平均F値は48%と分類は行えていない。SVMでは、被験者全体の平均F値は50%と分類は行えていない。

次に、右腕の指先から手首まで以外の骨格座標を取得した場合の評価を行った。CNNでは、被験者全体の平均F値は50%と分類は行えていない。SVMでは、被験者全体の平均F値は46%と分類は行えていないが、被験者によっては71%、81%となっており、分類できている。よって、右腕の肘に特徴があらわれていることがわかった。

最後に、Wilcoxon の符合付順位和検定を用いて、取得する骨格座標を変更した際に有意かどうかを確認した。まず、元データ以外の7つの代表値が、元データの代表値と有意差がないことを帰無仮説とし、有意差があることを対立仮説として、有意水準0.05で両側仮説検定を行った。その結果、元データを0.7倍や1.3倍にした線形補間データについて帰無仮説が採択され、元データの代表値との間に有意差は認められなかった。元データをスライドさせたデータ、指先から肩までのデータ、指先から肩まで以外のデータ、指先から肘まで以外のデータ、指先から手首まで以外のデータについては帰無仮説が棄却され、元データの代表値に対して有意差があった。

次に、指先から肩までのデータ以外の3つの代表値が、指先から肩までのデータの代表値と有意差がないことを帰無仮説とし、有意差があることを対立仮説として、有意水準0.05で両側仮説検定を行った。その結果、指先から肩まで以外のデータ、指先から肘まで以外のデータ、指先から手首まで以外のデータについて帰無仮説が棄却され、指先から肩までのデータの代表値に対して有意差があった。

次に、指先から肩まで以外のデータ以外の2つの代表値が、指先から肩まで以外のデータの代表値と有意差がないことを帰無仮説とし、有意差があることを対立仮説として、有意水準0.05で両側仮説検定を行った。その結果、指先から肘まで以外のデータ、指先から手首まで以外のデータについて帰無仮説が棄却され、指先から肩まで以外のデータの代表値に対して有意差があった。

指先から肘まで以外のデータ以外の1つの代表値が、指先から肘まで以外のデータの代表値と有意差がないことを帰無仮説とし、有意差があることを対立仮説として、有意水準0.05で両側仮説検定を行った。その結果、指先から手首まで以外のデータについて帰無仮説が採択され、指先から肘まで以外のデータの代表値との間に有意差は認められなかった。

本論文における実験の結果、CNNよりSVMのほうがうまく二値分類できていることがわかった。なお、右腕の指先から肩までの骨格座標を取得した場合には、平均73%、被験者によっては97%の精度で区別ができ、ピッキングおよび鍵開け動作を行っている右腕に特徴があらわれていた。本システムでは、画像や映像をセンタにデータ転送する必要がなく、家庭内で行動の判別が可能である。そのため、本提案手法は、家庭の防犯対策に有用である。