

氏 名	KEYA DAS TILOTTOMA
博士の専攻分野の名称	博士（学術）
学位記号番号	博理工甲第1170号
学位授与年月日	令和2年3月23日
学位授与の条件	学位規則第4条第1項該当
学位論文題目	Detecting of internal emotions using video cameras (ビデオカメラを用いた内面的感情の認識)
論文審査委員	委員長 教 授 久野 義徳 委 員 准 教 授 小室 孝 委 員 准 教 授 小林 貴訓 委 員 教 授 島村 徹也

論文の内容の要旨

Automatic recognition of emotions in humans is a challenging task with many applications such as human robot interaction, movie marketing, and more. Thus, there has been much work on systems that identify emotional states. Human emotion recognition using facial expressions is a common approach and there are many researchers working in this direction. However, there are times when facial expressions can either be faked or hidden. That is, one's "apparent emotions" from say, facial expressions may not be a reflection of one's genuine "internal emotions". Thus, other modalities such as physiological responses should be investigated for detecting and recognizing a person's internal emotions. This thesis aims to build a practical system for detecting such emotions by observing physiological responses. We focus on sensing physiological changes through visual means using only conventional cameras, as this would not require specialized equipment.

As mentioned earlier, the computer vision community has made many advancements in apparent emotion recognition through facial expressions. On the other hand, the psychophysiology community has conducted several studies on detecting and recognizing internal emotions using different physiological channels. Recognition of emotion has been done using many physiological signs such as heart rate change, eye movements, eye blinks, change in skin conductance, and change of skin temperature. Although many different physiological signs are used, many researchers find that cardiac activity is useful for emotion recognition. As a result, this thesis explores the use of cardiac activity for emotion recognition. However, most past studies use electrocardiography (ECG) for reading cardiac activity. ECGs are effective but have their limitations due to the need for attached sensors and higher cost. To realize practical application systems for many real-world settings, we also need a method that can sense cardiac activity but does not require any wearable attachments or devices. Fortunately, in recent years, remote photo plethysmography (PPG) algorithms have received attention. These techniques allow one to read cardiac activity such as heart rate (HR) from conventional cameras by typically observing small changes in skin color over time. This thesis aims to use the sensed cardiac activity from remote PPG to detect and recognize emotions. Since remote PPG has been shown to work with conventional cameras,

our proposed approach has the benefit that cameras such as webcams, surveillance cameras, and cellphone cameras could be used. With the ability to see cardiac activity without contact sensors, we present a convenient system for detecting internal emotions. Like in the psychophysiology literature, we chose to evaluate our approach by recognizing emotional reactions to emotionally stimulating videos such as horror and comedy clips.

In the first phase of our work, we showed video contents to human subjects and collected HR data using an attached sensor (Fitbit) for three emotional states (normal resting, funny, and horror). We then confirmed that the average HR in emotionally stimulated states exhibits a statistically significant increase ($p\text{-value} < 0.01$) from that in normal resting states. We then estimated HRs using remote PPG to analyze videos of the subjects' faces and found them to be strongly correlated (0.9) with the wearable sensor ground truth data. The increase in average HR from normal resting to emotionally stimulated states was also similar to the increase observed in the case of the wearable sensor.

The first phase of our work showed that HRs could be used to detect changes in emotional state but did not explore the recognition of what kinds of emotions were present. In the next phase of our work, we investigated the feasibility of using cardiac pulse signals for recognizing different emotional states (joy vs. fear) and compared the results with those based on facial expressions. Specifically, we used the Open Face facial landmark tracker to estimate the average facial action unit intensities for each subject on the 30 second segments in both the comedy and horror cases. In all, 17 facial action units were used. We also modified our remote PPG algorithm and showed that cardiac pulse signals, as opposed to only HR can be estimated. We then used linear Support Vector Machines (SVMs) to evaluate the feasibility of recognizing whether the subject watched a horror or comedy clip. That is, we used SVMs to test the leave-one-out cross validation (LOOCV) performance of facial action units in emotion recognition and repeated the same test for the case of cardiac pulse signals estimated from remote PPG. We found that with principal component analysis (PCA) preprocessing, facial action units resulted in a LOOCV accuracy of 78.8% while the cardiac pulse signals resulted in a LOOCV accuracy of 67.3%.

We have presented an approach to emotion recognition based on physiological responses rather than facial expressions. As a result, the emotions that we detect cannot be easily faked. In addition, these cardiac pulse signals are entirely estimated from videos captured by a conventional RGB camera, so no special equipment is required in our setup. Essentially, we have a system that operates completely using only computer vision techniques. A drawback of the current study is that our dataset did not have the data for human subjects that intentionally tried to hide or fake their emotions, thus it was not ideal for our tests. The results suggest that in situations where human subjects have no reason to hide their emotions, facial expressions are reliable. However, we show that we can achieve good accuracy even with a naïve approach like taking the estimated cardiac pulse signals, performing dimensionality reduction using PCA, and then learning via linear SVM.

To our knowledge, we are one of the first to bridge the gap between computer vision and psychophysiology through presentation of a promising system for visual detection of internal emotions. This thesis paves the way for future research into detecting and recognizing internal emotions.

論文の審査結果の要旨

当論文審査委員会は、当該論文の発表会を令和2年2月4日に公開で開催し、詳細な質問を行い論文内容の審査を行った。その論文発表を含む学位論文の審査の結果、本提出論文を博士（学術）の学位論文として合格と判定した。以下に審査結果の要約を示す。

本提出論文は、ビデオカメラの映像から心拍関連の情報を取得し、そこから当人の真の内面の感情を推定する技術に関して検討したものである。感情の状況が分かれば、それに応じて適切な対応をするコンピュータシステムやロボットが実現できると期待される。感情は一般に表情に現れるので、表情からの感情認識がコンピュータビジョンの分野では多く研究されてきた。しかし、感情の状況が変化しても表情に現れない場合も多い。一方、生理心理学では、心電図、脳波、体温、皮膚電位などの生理学的指標から感情の状況を推定する方法が検討されてきた。生理学的指標は表情のように意識的に可変はできないので真の感情状況を示すと考えられるが、計測には身体に装着するセンサ等が必要であり、コンピュータやロボットとのインタラクションの応用には使いにくい。しかし、近年、ビデオカメラの性能が向上し、皮膚の微小な色の変化が検出できるようになったことにより、画像から生理学的指標の一つである心拍数やそれに関連する情報を取得する方法が提案されるようになってきた。これによれば、身体に何も装着する必要なく、カメラでユーザを見ることにより、ユーザの真の内面的感情を推定することができると考えられる。本論文は、この問題について基礎的な検討を行ったものである。

本論文は6章からなる。まず、第1章では、上で述べたような本研究の背景、研究目的について述べている。そして、ビデオ画像からの内面的感情の推定における検討課題を明らかにしている。

第2章では、関連研究について述べている。画像からの表情やその他の非言語行動からの感情推定に加え、生理学的指標に基づく方法を調査している。さらに、感情推定の結果を利用したコンピュータやロボットと人間のインタラクションに関する関連研究についても調べている。

第3章では、ビデオ画像からの心拍数計測により内面的感情の推定が可能かどうか検討している。ビデオ画像からの心拍数計測については、所属研究室で開発された、多数の小領域対の輝度変化から推定する方法を用いている。実験では、被験者にホラー映画やコメディーを見せたときの被験者の顔画像を撮影する。被験者には装着型の心拍数センサも付けてもらう。はじめに、装着型のセンサとビデオ画像からの心拍数の計測結果を比較して、ビデオ画像からの計測が正確であることを確認した。各種の映像を見せた場合の結果から、映像により感情状況が変化したと思われる場合に心拍数が変化することを確認し、ビデオ画像からの心拍数計測により感情の変化が検出できることを示した。

第4章では、感情の変化だけでなくどのような感情かの認識の可能性について検討している。そのためには心拍数だけでは情報が不足するので、ビデオ画像から心臓の拍動に対応する信号を求めて用いる。ここでは、より正確な信号を得るために、第3章で用いた方法を改良した。第3章で用いた方法では顔の上にランダムに設定した小正方形領域を追跡して、その領域内の平均輝度を利用して心拍数を推定した。同じ部分の輝度変化をさらに正確にとらえるため、顔の上の特徴点を求め、特徴点を結ぶ3角形領域の平均輝度を用いるように改良した。これを用いて、毎秒30フレーム30秒間の信号の900点の輝度データを特徴ベクトルとして、それをSVM (Support Vector Machine) で識別する。これにより、ホラー映画を見たときとコメディーを見たときの心拍信号の識別が可能かを実験した。比較のために、表情認識による感情推定法も用意し

た。その結果、心拍信号を用いた方法で67%の認識率が得られた。表情認識による方法では79%であった。コメディーを見たときは笑う場合が多いため、表情認識の方が認識率は高いが、表情の変化があまり見られない場合や、さらに表情は意図的に作ることもできることを考えると、心拍信号によりごまかすことのできない内面的感情をある程度識別できることが示された。

第5章では、ビデオ画像の撮影に加えて、各種の装着型センサも利用してもらって、各種のビデオ映像を見た場合について計測を行っている。これにより、画像からの心拍計測と他の方法による計測を比較している。また、この画像とセンサデータは今後の感情推定の研究のための有用なデータセットとなる。

最後に第6章で、全体を総括し、今後に残された課題を議論している。

各種の感情の状況のデータを集めるのは難しい。今回は、ホラー映画とコメディーを見てもらうことにより、感情の変化が引き起こされると仮定して検討を行った。まだかなり限定された状況での検討だが、ビデオ画像からの心拍信号から感情の変化の検出や種類の識別の可能性が示された。認識精度の向上や、さらに他の感情への対応など検討課題はあるが、人間の気持ちに応じて対応できるコンピュータやロボットへの応用可能性は示すことができたと評価できる。

以上のように、本論文の内容は、学術的に意義のある研究であると判断できる。よって、当学位論文審査委員会は、本論文を博士（学術）の学位論文として合格と判定した。