様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科研費
KAKENHI

科学研究費助成事業　　研究成果報告書

令和　　元 年　　６ 月 １４ 日現在

機関番号：１２４０１
研究種目：若手研究(B)
研究期間：2017～2018
課題番号：１７Ｋ１２７０９

研究課題名（和文）Real World Heart Rate Estimation from Video

研究課題名（英文）Real World Heart Rate Estimation from Video

研究代表者
Ｌａｍ　Ａｎｔｏｎｙ（Lam, Antony）

埼玉大学・理工学研究科・助教

研究者番号：５０７４４１２４

交付決定額（研究期間全体）：（直接経費）　　3,100,000円

研究成果の概要（和文）：本研究課題では顔のビデオ画像から人間の目には見えない皮膚の色の微小変化を捕らえ心拍数を推定するアルゴリズムの開発を目指した。我々が以前に提案した実環境での照明変化に頑健な方法をもとに、さらに頭部の回転に対しても対応できるようにして性能を向上した。これは、頭部の様々な回転に対して適応的に変形する顔面上の3角形局所領域を追跡することにより達成した。この方法はウェブカメラに対しても動作することを確認した。さらに、他の様々な応用に利用できるように精度は多少劣るが高速なアルゴリズムも開発した。これは、心臓の活動により色の変化が顕著に現れる顔の特定の領域を考慮することにより実現した。

研究成果の学術的意義や社会的意義
Our work is one of the few in the literature that addressed the difficult challenges of estimating cardiac activity from videos with changing illumination and motion in the real-world. This algorithm can also be used for applications such as emotion recognition and human computer interaction.

研究成果の概要（英文）：Our aim in this project was to develop an algorithm that can take a video feed of someone's face and through reading subtle changes in skin color (imperceptible to the human eye), estimate the person's heart rate. We took our previous work that was robust to lighting changes in real-world environments and improved it by adding robustness to out-of-pose head rotations. This was achieved by tracking triangular local regions of the face where the regions could adaptively change shape to fit many 3D rotations of the head. The tracked triangular local regions of the face were then used to perform a robust estimate of cardiac activity from the subtle color changes of the skin. We found this technique works even with RGB webcams. We also developed a variant of the algorithm that is faster but less accurate so it could be used for different applications. This variant worked by considering specific face regions with strong cardiac-caused color changes in skin.

研究分野：Computer Vision

キーワード：video ppg　computer vision　heart rate　cardiac estimation　remote measurement

１．研究開始当初の背景 (Background of the Beginning of the Research)

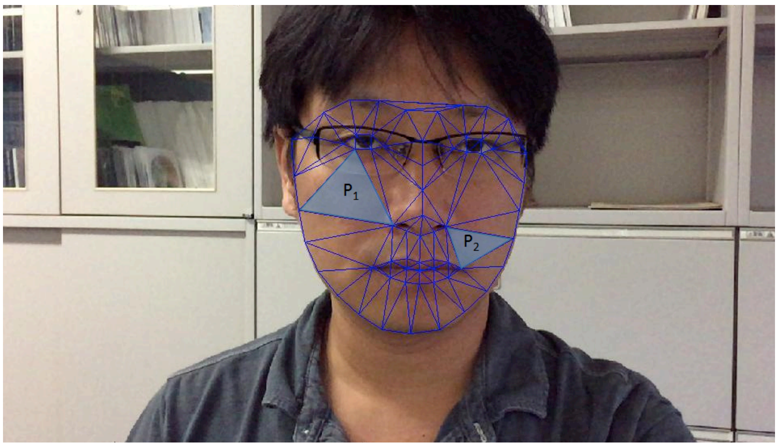
In 2008, Verkruysse, et al. reported in their paper, "Remote Plethysmographic Imaging Using Ambient Light", that conventional RGB cameras can be used to sense very weak color changes in human skin that correspond to cardiac activity. However, their work required the environment's illumination to remain constant and their human subjects could not move. In 2015, we developed an algorithm that could analyze multiple local face regions from RGB video and combine all results into a single robust estimate of a given person's heart rate (HR) even in the presence of changing illumination (see "Towards Estimating Heart Rates from Video Under Low Light" by Lam and Kuno). This work was published at the International Conference on Computer Vision. A drawback of our 2015 algorithm is that it is sensitive to motion (especially for 3D head rotations) and it is computationally expensive.

２．研究の目的 (Purpose of Research)

In this project, our aim has been to effectively estimate HR from RGB videos captured under real-world settings. This means retaining the illumination robust capabilities of our original algorithm while adding robustness to realistic motions one would expect from real-life human users. In addition, we also aimed to explore potential applications of remote cardiac sensing through RGB cameras in areas such as Emotion Recognition, Human Computer Interaction, and Human Robot Interaction. The idea being that if we can sense physiological changes (e.g. cardiac activity) without any attached sensors, this would provide us with a powerful ability to sense a user's state of mind and allow for technology to better serve and interact with the user.

３．研究の方法 (Research Method)

One of the first steps in our research was to build a good dataset of test videos. This is because we found that although well-known datasets such as the MAHNOB-HCI (see "A Multimodal Database for Affect Recognition and Implicit Tagging" by Soleymani et al.), contain a large number of very high quality video and physiological readings for people, most of the human subjects are static. As a result, we conducted some tests with more natural motions by recording our own videos. We also collected some YouTube videos so that we could test our algorithm on real-world videos.



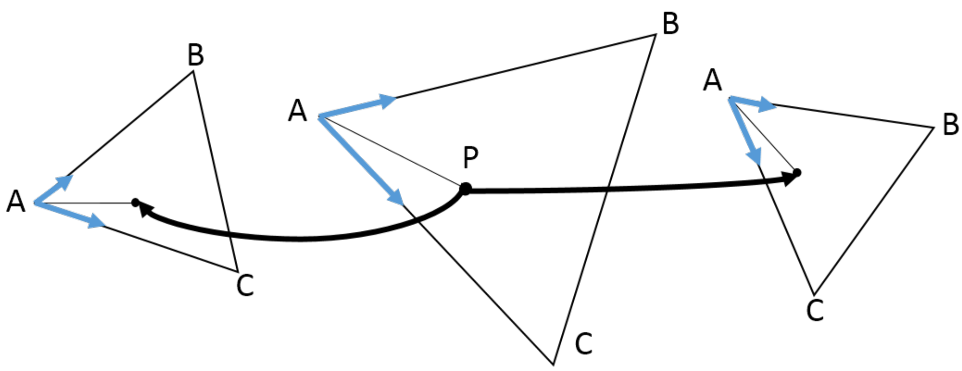


Fig. 1. Local triangular face regions (top) and Barycentric Coordinates for tracking triangles as they deform across frames (bottom).

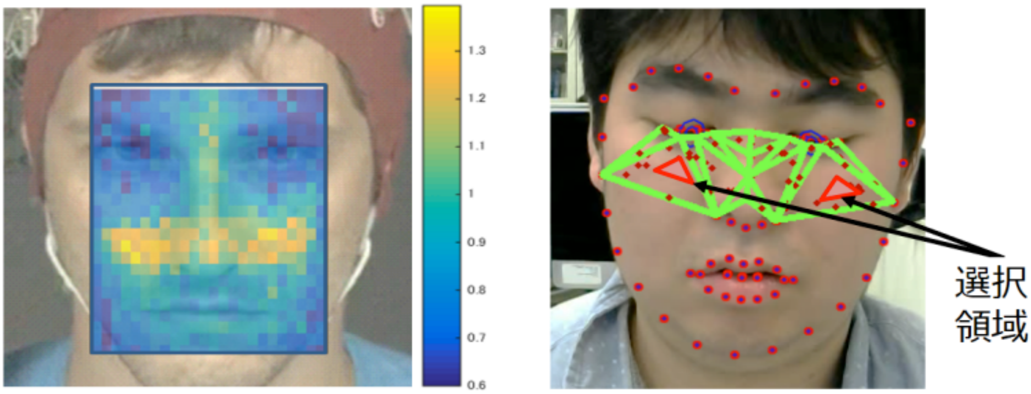


Fig. 2. Variant of our algorithm that uses face regions with the strongest skin color changes due to cardiac activity. Heatmap of strong cardiac signals (left) and selected key face regions for HR estimation (right). This approach can be used for many applications.

In the next step, we explored ways to overcome our 2015 algorithm's main limitation concerning 3D head rotations. We devised a Delaunay Triangulation patch-based approach to deal with 3D head rotations. Specifically, given a video of a person's face, we first detected and tracked facial landmarks. We then used Delaunay Triangulation to compute triangular local regions on the face. Then, using Barycentric Coordinates, we devised an efficient means for tracking these triangular regions by using the known tracks of the facial landmarks of the face. (See Fig. 1 for an illustration of these processes.) This algorithm was tested on both the motion videos we collected using our own cameras and on YouTube videos.

At the same time, we also considered some applications where our algorithm could be used in more constrained settings. That is, we do not always need the algorithm to handle any kind of unconstrained motion. For this case, we assumed the motions of the human user would be more controlled and aimed to develop a variant of our algorithm that would be more focused on speed. To achieve this, we conducted experiments to determine which parts of the human face tend to have the strongest color change signals due to cardiac activity. (See. Fig. 2.)

We also explored some potential applications. In particular we have found that the estimated pulse signal (not the HR) from our algorithm can be used to classify emotional states. In addition, we are also considering other applications in Human Computer Interaction and Human Robot Interaction.

４．研究成果 (Research Results)

One of our challenges was to improve over our older algorithm from 2015 so that it could handle real-life head motions such as rotations. To test this, we devised the triangulation and triangle deformation tracking approach illustrated in Fig. 1. We first tested on 18 self-made videos captured using a Logicool C920 Webcam with realistic head motions. We found that HRs were estimated with a Mean Absolute Error (MAE) of 6.9 BPM. This was promising but we wanted to test on even more realistic videos such as vlogs from YouTube so we also tested the viability of HR estimation algorithm on YouTube. We uploaded our 18 self-made videos to YouTube and downloaded them in YouTube's format 135. After testing our algorithm on the videos downloaded from YouTube, the MAE was 9.9 BPM. Thus we see that video-processing artifacts can increase error but the HR can still be estimated relatively well.

In testing on real vlogs, we would not have the ground truth HRs of the vloggers. However, we can indirectly observe HR changes due to emotional changes. (Our study [雑誌論文 1] shows that emotional stimulation can increase HRs that are detectable by video.) Thus we found YouTube reaction vlogs (Fig. 3), which consist of real vloggers watching the same horror movie trailer. Our hypothesis was that the average HR of all the vloggers watching the same horror movie trailer should show HR increases during the scariest parts of the movie trailer (Figs. 4 and 5).

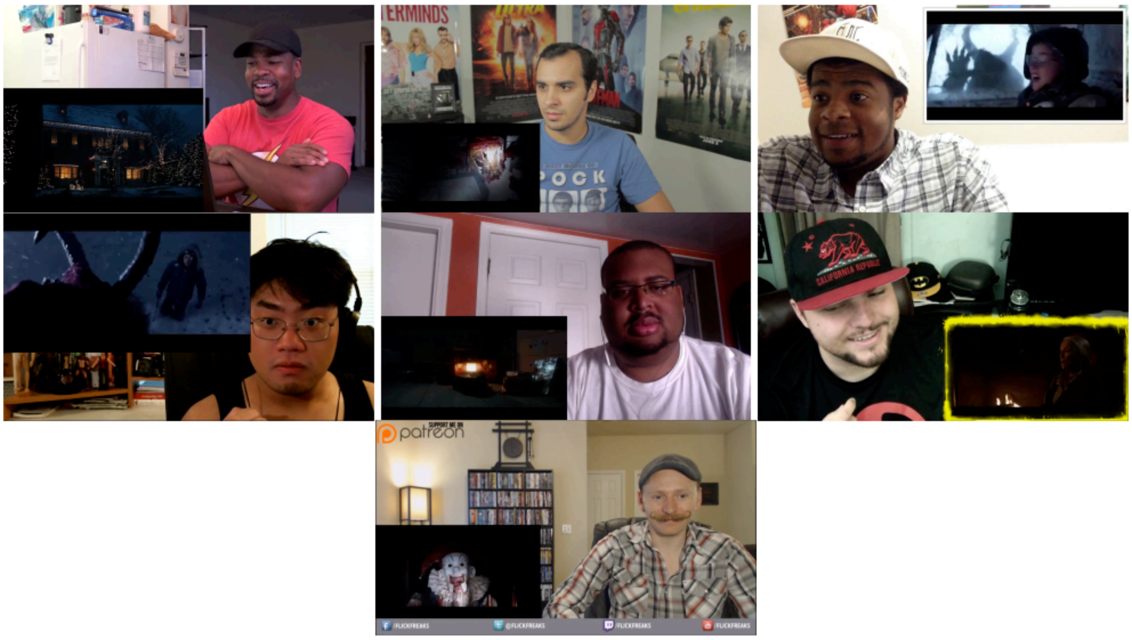


Fig. 3. YouTube reaction videos to the same horror movie trailer. We applied our algorithm to estimate HR changes to each vlogger.

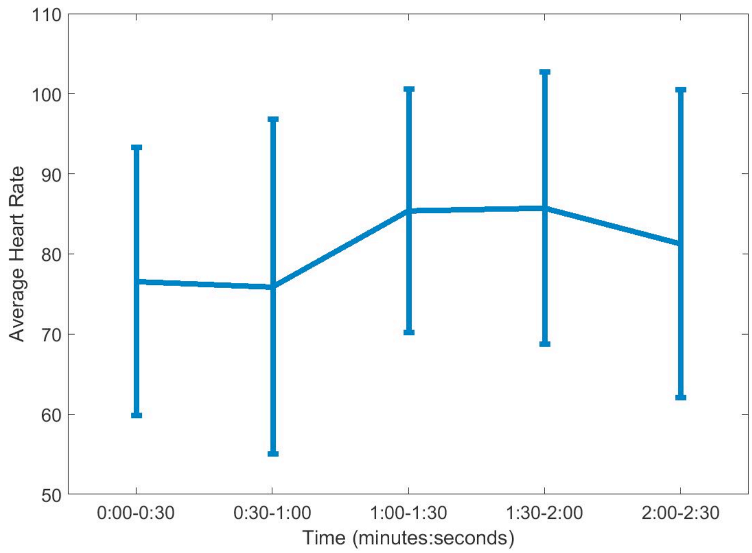


Fig. 4. Average HR in BPM of the YouTube vloggers with standard deviations. The scariest parts of the movie trailer caused the average HR to increase.

While we could not determine if the HR estimates for each YouTube vlogger was correct, we did see that the average HR increases of the vloggers was as expected. This provides an indication that our algorithm does work even in the difficult case of some YouTube vlogs. We hope to conduct a much larger-scale study in this direction in the future. (Note that we cannot compare the improvements in HR estimates over our 2015 algorithm. This is because our 2015 algorithm cannot handle frequent 3D head rotations.)

A drawback of our algorithm is still its high computational cost. But for many applications, we may be able to assume that the frontal face is highly visible and motion is more limited. Thus we developed the variant that uses key face regions (Fig. 2). This algorithm resulted in improved processing speed while maintaining and even improving accuracy in some cases (Table 1).

Table 1. Comparison Between Video HR Estimation Algorithms

Algorithm	Number of Local Regions Processed	Processing Time (Seconds)	Mean Absolute Error (BPM)	Mean Squared Error (BPM)	Percent <5 BPM Absolute Error
Our 2015 Algorithm	500	271 secs.	6.03	10.99	70.7%
Optimized Key Region Selction (Fig. 2)	100	87 secs.	5.31	9.8	72.6%

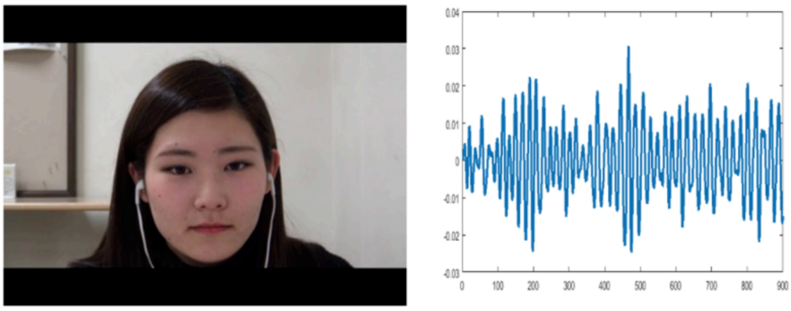


Fig. 5. Human Subject with Estimated Cardiac Pulse from Video

Another direction we explored was in applications. We considered how estimated HR from video could be used to classify emotions in people. The motivation for this is that facial expressions do not always reveal a person's true emotional state because some people are not expressive or they may be trying to mask their emotions. So we conducted our study on emotion recognition by building our own dataset of human reactions to emotionally stimulating videos. Our dataset consists of 26 subjects that watched comedy and horror movie clips for a total of 10 minutes (5 minutes per clip). We then estimated their HRs from video but found HRs alone could not be used to differentiate reactions to watching comedy versus horror video clips. Fortunately, our algorithm can also be used to estimate the full cardiac pulse signal as opposed to only the HR (Fig. 5). (Please see our paper [雑誌論文 2] for details). Using the full cardiac pulse signal estimated from video, we then found that simply using a combination of PCA and linear SVM learning, we could differentiate horror versus comedy reactions (without using facial expressions) with an accuracy of 67.3%. We are now working to improve the emotion recognition accuracy of our system and also exploring potential applications in Human Computer Interaction and Human Robot Interaction.

５．主な発表論文等 (Main Presentations and Papers)

〔雑誌論文〕（計 3 件）[Journal Articles]

1. Keya Das, <u>Antony Lam</u>, Hisato Fukuda, Yoshinori Kobayashi, and Yoshinori Kuno, Classification of Emotions from Video Based Cardiac Pulse Estimation, International Conference on Intelligent Computing, Peer Reviewed, Vol. 10956, 2018, 296-305.
2. Keya Das, Sarwar Ali, Kouyou Otsu, Hisato Fukuda, <u>Antony Lam</u>, Yoshinori Kobayashi, Yoshinori Kuno, Detecting Inner Emotions from Video Based Heart Rate Sensing, International Conference on Intelligent Computing, Peer Reviewed, Vol. 10363, 2017, 48-57.

〔学会発表〕（計 4 件）[Conference presentations]

1. 大津耕陽, Keya Das，福田悠人, <u>Antony Lam</u>，小林貴訓，久野義徳，映像解析に基づく頑健・高速な心拍数計測手法，第 24 回画像センシングシンポジウム (SSII2018) 予稿集，DS2，2018.
2. <u>Antony Lam</u>, Kouyou Otsu, Keya Das, Yoshinori Kuno, Towards Taking Pulses Over YouTube to Determine Interest in Video Content, International Workshop on Frontiers of Computer Vision, Peer Reviewed, 2018.
3. Keya Das, Kouyou Otsu, <u>Antony Lam</u>, Yoshinori Kobayashi, Yoshinori Kuno, Towards Detecting the Inner Emotions of Multiple People，第 99 回パターン 計測部会研究会，2017.
4. 大津耕陽，倉橋知己，Tilottoma Das，福田悠人，<u>Antony Lam</u>，小林貴訓，久野義徳，環境変化に頑健なビデオ映像による心拍数計測手法，第 23 回画像センシングシンポジウム (SSII2017) 予稿集 IS2 巻, 2017, 20.

〔その他〕
ホームページ等
http://yankee.cv.ics.saitama-u.ac.jp/~kunolab/antonylam/
https://scholar.google.com/citations?user=N39EWuEAAAAJ&hl=en