A Smartphone Tool for Evaluating Cardiopulmonary Resuscitation (CPR) Delivery

Gavin Corkery and Kenneth Dawson-Howe
School of Computer Science and Statistics, Trinity College, University of Dublin, Ireland

Keywords: Video Processing, Machine Vision, Optical Flow, Cardiopulmonary Resuscitation.

Abstract: This paper presents a prototype smartphone application to aid with the delivery of cardiopulmonary resuscitation (CPR). The person giving CPR is viewed from the side and both compressions and breaths are identified primarily using optical flow. This allows the system to provide near real time feedback on the chest compression rate (CCR) and on the timing of breaths (which affects the Chest Compression Fraction (CCF)). The system is evaluated on over 25 minutes of video of 6 different participants delivering CPR to a test dummy. A quantitative evaluation is presented which shows that the system recognised 99% of compressions and all of the breaths (although two false positive breaths were classified). It computed the CCF to within 1%.

1 INTRODUCTION

The use of cardiopulmonary resuscitation (CPR) has been shown to increase survival levels from 5.2% to 7.8% in the case of normal CPR and to 13.3% in the case of Compression-only CPR (COCPR) (Bobrow et al., 2010). In the same study they found that CPR or COCPR was only administered in around one third (34.3%) of the out-of-hospital cardiac arrest incidents. However it seems that this rate can be significantly increased (to around 60%) when proposed by the emergency services telephone dispatcher (Dami et al., 2010). In addition, it appears that if a defibrillator (AED) is available the survival rate increases significantly. In a study by Iwani et al. (Iwami et al., 2012) considering only cases where an public access AED was used, the survival rate for normal CPR was 32.9% and for COCPR was 40.7%.

1.1 CPR Guidelines

The (American Heart Association (AHA)) recommendations for CPR are currently that chest compressions (CC) should be applied at a rate (CCR) of 100-120 per minute to a depth (CCD) of 5-6cm allowing full recoil (decompression of the chest) after each compression (AHA, 2015). The two hands should be placed (one on top of the other) with the heel of the lower hand in the centre of the chest on the lower half of the breastbone (sternum), and compressions should be done firmly and smoothly pressing downward while keeping the arms straight. For a layperson who is trained in giving rescue breaths it is recommended that the chest compressions should be alternated with rescue breaths in a ratio of 30 compressions to 2 breaths. Each breath should be delivered over one second causing the chest to rise, and the chest should fall again between breaths. The period taken for the breaths reduces the chest compression fraction (CCF - i.e. the portion of time during which compressions are performed) which should be as high as possible with a target of at least 60%.

1.2 Existing Systems for Aiding/Evaluating CPR

The AHA guidelines state that it “may be reasonable to use audiovisual feedback devices during CPR for real-time optimization of CPR performance”, but notes that “studies to date have not demonstrated a significant improvement in favorable neurologic outcome or survival to hospital discharge with the use of CPR feedback devices during actual cardiac arrest events” (AHA, 2015). In recent years the range of technology available to provide CPR quality feedback has significantly expanded so there is some potential for this feedback to improve positive outcomes. In addition such technologies provide methods for improving CPR training.
In lab/training scenarios, transmitter-receiver pairs have been used for evaluating chest compression depth (Kim et al., 2017; Kang et al., 2010) with high precision. Also in a lab environments evaluation (of CCR, CCD and CCF) has been done using RGB-D sensors (Higashi et al., 2017; Loconsole et al., 2016) and balance boards (Hayashi and Minazuki, 2017; Higashi et al., 2017; Ferreira et al., 2017) which can give a measure of the force direction as well as the depth. These preceding technologies are only appropriate for use in lab/training environments but a number of new technologies have potential in the field.

Every AED has sensors built into it and these have been used to measure both CCR and CCF (Torney et al., 2016; Gonzalez-Otero et al., 2012). Accelerometers have been used independently (Yamamoto and Ohmura, 2015; de Gauna et al., 2015), on smartwatches (Ahn et al., 2016), or smartphones (Song et al., 2015; Amemiya and Maeda, 2013) to evaluate CCD and CCR. In addition smartphone cameras have been used to evaluate CCR based on a view looking upwards at the person applying CPR (Meinich-Bache et al., 2017; Frisch et al., 2014), where the smartphone is lying flat on the ground. There are hundreds of smartphone applications which aid in the training and delivery of CPR (Ahn et al., 2016) although most of these tools do not provide feedback regarding the quality of the CPR being given. For CPR tools which do provide CPR quality feedback, this is done as audio feedback, typically in the form of a metronome (Amemiya and Maeda, 2013; Loconsole et al., 2016) or messages such as “Push Faster”, “Good Speed”, “Push Slower”) (Torney et al., 2016) and as video feedback, typically a colour indication of CCR quality (Ahn et al., 2017; Amemiya and Maeda, 2013).

Only one piece of research (that we could locate) considers the posture of the person performing CPR in terms of keeping the arms straight (Higashi et al., 2017). No research could be located which consider the position of the hands on the chest of the patient.

In addition Panicker et al. presented work identifying CPR scenes in medical. simulation videos (Panicker et al., 2015)

1.3 System Concept

The research presented in this paper describes the creation and evaluation of the first version of a smartphone application intended to aid in the delivery of CPR. The concept of the system is that it should assess the CCR (and the breathing phases) based on the analysis of a video of a person giving CPR taken on a smartphone. This evaluation must happen as close to real-time as possible and provide feedback on the CCR so that the person can increase or decrease the compression rate appropriately.

2 SYSTEM

To recognise both the compressions and the breaths we rely on dense optical flow (as implemented in OpenCV); See Figures 1 and 2.

In order to eliminate some noise, the amount of movement from frame to frame must meet a certain threshold before it is considered. This reduces the impact of very small movements on the algorithm. Examining video footage showed a great amount of movement in the compressions, so a small threshold does not limit the recognition of this movement. The threshold was chosen to be quite small, requiring at least 0.5 pixels of movement between frames. This recognises the moving region of interest while reducing a lot of noise.

In order to reduce noise from the movement of background objects, a model of the moving and non-moving regions of the scene is maintained. As the CPR session progresses, the regions with a great amount of historical movement are given a greater weighting over those that have been historically static, meaning that background movement will cause fewer false positives. The weighting is updated at each frame for each pixel by the amount of movement from the previous frame, multiplied by a small learning rate (experimentally chosen to be 0.005). After several compressions, there will be a clear model of
the moving regions of the image, which include the head, chest and arms of the person giving CPR. The weighting is not updated during the breathing phase, as the movement model is focused solely on the movement during compressions.

2.1 Recognising Compressions

Having a weighted motion field for each frame forms the basis for recognising chest compressions. Through analysis of test videos, it is clear that a chest compression can be described by two phases: a strong upward motion phase and a strong downward motion phase (See Figure 1). A compression is deemed to be found when strong upward movement occurs after a strong downward movement. The upward movement must be within a small window of time after the downward movement.

Compressions are characterised by both strong upward and downward movement. To account for noisy scenes, the dominant movement must also be several times (the ratio chosen was 3) greater than the movement in the other direction. This ratio is useful in a scene with a lot of background movement. There is also a threshold for the amount of the scene that must be moving. This was chosen to be 10% to ensure that the movement takes up at least a small portion of the scene.

In order to reduce false positives, the timing of the upward movement in relation to the downward movement is also measured. For a compression to be recognised, the upward movement must take place no longer than 1 second after the downward movement. This value was chosen to reduce false positives, however it may cause difficulties with recognising extremely slow compressions.

2.2 Recognising Breaths

The recognition of breaths uses a similar technique to the recognition of compressions. The recognition of a breath is split into two phases: the downward lateral movement and the upward movement at the end of the breath (See Figure 2).

The lateral movement is recognised by using a threshold based on the mean lateral movement during compressions. The total movement to the left or right must be 5 times greater than the mean movement in that direction. The lateral movement must also be at least a third of the downward movement. This movement must take place in at least 3 consecutive frames to be counted as the start of a breath. This is to ensure that the movement is actually taking place. The upward movement at the end of the breath is recognised in the same way that the end of a compression is recognised. This also acts as a mechanism for recovering from error if a breath was falsely identified.

The overall algorithm for processing individual frames is illustrated in Figure 3.

2.3 Calculating CCR and CCF

In order to provide useful feedback, the application calculates the CCR while performing compressions, and the CCF when the CPR session has concluded.

The CCR is calculated using a weighted learning approach. For each compression, the calculated rate is computed as 36000 (the number of milliseconds in a minute) divided by the number of milliseconds between compressions. This means that if there is half a second (500 milliseconds) between two compressions, the rate is 120 compressions per minute. The
"time" of a compression is defined as the time of the first strong upward movement after downward movement.

For the first 5 compressions, the impact of each new compression is 50% of the overall rate. This allows for a quick establishment of the rate at the start of compressions. After the first 5 compressions, each new compression makes up 15% of the overall rate. The reason for this lower rate is to minimise the user’s oscillation between slow and fast compressions due to negative feedback causing vast overcorrections.

The CCF is calculated using the following formula:

\[
CCF = \frac{(Total\ Time - Interrupted\ Time)}{Total\ Time}
\]

In this case, the interrupted time refers to the sum of the periods of time which were at least 2 seconds long and did not contain any compressions. This will include time taken for breaths.

3 MOBILE APPLICATION

The implementation of the aforementioned techniques is realised through a smartphone application. The landing page of the application provides the user with a view of the current scene through either the front or rear camera, depending on user selection. In training/practice situations, the smartphone should be placed at a close distance from the CPR dummy, with the CPR dummy lying between the smartphone and the rescuer. Since different smartphone cameras will have different fields of view, a precise optimal distance from the phone cannot be recommended. Ideally, the rescuer and dummy should both be clearly visible and fill a majority of the camera view.

3.1 Realising Acceptable Frame Rate

The most important aspect to consider in an application processing live video is the resulting frame rate after processing each frame, which can be quite expensive. Some of the videos were of a very high resolution (1920x1080) and this resulted in a frame rate of approximately 1 frame per second (fps). To handle this, each image frame is resized to a 216x216 pixels frame, which is over 25 times smaller. This gives a massive performance boost without degradation to the computer vision algorithm. The resulting performance increase affords a frame rate of approximately 15fps, which we found to be sufficient to provide high quality feedback to the user. In addition, for speed purposes, the vision algorithm does not analyse every pixel, but samples at a constant rate. This is due to the observation that the moving area will be a large, mainly contiguous area. Analysing every pixel is therefore unnecessary so we analyse every 16th pixel in each row or column is analysed, leading to \( \frac{1}{16} \) of movement to analyse. This spacing works well for all tested distances, but may be too wide if the rescuer is extremely far from the camera. The assumption, however, is that the rescuer will never be so far away from the camera for this to be an issue.

3.2 Methods of Feedback

As the person giving CPR would be focused on the task, it is important that feedback is clear and does not require careful analysis of the phone screen. The visual feedback is as obvious and legible as possible, while audio feedback is used to provide evaluation without having to focus on the smartphone screen. The information needed to provide metrics are processed after the previous camera frame has been processed, ensuring that feedback is always as up-to-date as possible.

3.2.1 Visual Feedback

The primary measurable metric of CPR by the application is the chest compression rate. When performing CPR compressions, this is the aspect that should be the main focus of the administrator’s attention. The region at the top of the screen is used to provide an easily-readable indicator of their current rate. The rate is shown as a large number, which is colour-coded to align with the recommended compression rate (See Figure 4). Rates between 100 and 120 compressions per minute are optimal, and are thus displayed with a green colour. Rates within 10 compressions per minute of being optimal are shown with orange, and rates further outside the optimal range are shown in red. The displayed rate is updated after every compression. This is done to ensure that any large rate changes can be rectified as soon as possible.

Visual feedback is also provided to tell the user to start breaths or to resume compressions (See Figure 5).

3.2.2 Audio Feedback

In cases where the person giving CPR may not be able to focus their attention on the phone screen, it is preferable to have another way of delivering feedback in a timely manner. This is achieved by using a speech synthesiser giving feedback with a clear voice.

The audio feedback is related to the same pertinent information dealt with by the visual feedback. During the compression phase, the application will tell
the user to keep going at their current rate if they are achieving an optimal rate. If they are slightly out of the optimal range they will be told to speed up slightly or slow down slightly, and if they are very far out of the range they will be told to speed up or slow down. The frequency at which the user receives feedback on their compression rate can be slow, medium or fast (feedback every 6, 4 or 2 seconds respectively).

Feedback is also provided for other important aspects. If the user chooses to perform the 30:2 cycles detailed above, they can choose to be notified to start artificial ventilation every 30 compressions. Along with this, they will be notified if their breaths have taken too long, and told to continue compressions. The audio aspect of this is important, as the user would not be focused on the phone screen if they are performing artificial ventilation.

### 3.3 Viewing Summaries

At the conclusion of the CPR session, the administrator is presented with a more in-depth analysis on the summary screen (See Figure 6).

The goal of the summary screen is to provide further insight into the user’s habits that may not be apparent from their CPR session. Examples of this could be that the user starts their compressions too slowly during each compression cycle, or that they are performing too few or too many compressions per cycle. Similarly to the visual feedback, the compression rate per cycle is colour-coded in the same rate ranges. The compression rate over time is graphed. If the user has improved their CPR quality significantly over time, this should be identifiable through analysis of current and past CPR sessions.

### 4 EVALUATION

The performance of the computer vision algorithm was measured on 11 pre-recorded videos of CPR being performed by ourselves and also by volunteers from the Dalkey First Responders Group. This footage consists of over 25 minutes of CPR with breathes being delivered to test dummies (8 videos with resolution 768x576 pixels 25 fps, and 3 videos with resolution 1920x1080 at 30 fps). To evaluate the performance of our system we created ground truth for this footage in terms of the frame numbers of the lowest point of each compression, and the presence of
breaths between sets of compressions.

4.1 Evaluation of Individual Compressions and Breaths

When performing the overall calculations for the test videos, the following formulae for accuracy, recall and precision are used (where TP is a True Positive: A compression or breath which occurred and was recognised by the system; FP is a False Positive: A compression or breath which was recognised by the system, which did not actually occur; FN is a False Negative: A compression or breath which did occur but was not recognised by the system):

\[
\text{Accuracy} = \frac{TP}{TP + FP + FN}
\]

Accuracy defines the proportion of times that the classification is correct.

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

Recall defines the proportion of breaths and compressions which are correctly identified.

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

Precision defines the proportion of times the classification is correct when it says a breath or compression has been identified.

4.2 Results

4.2.1 Compressions

The results for recognising compressions are very good. The vision algorithm has an accuracy, precision and recall of 0.99. The false positives mainly occur at the beginning and ending of the CPR session, where the rescuer is kneeling down and positioning their hands, or standing up and exiting the frame. This occurs due to the upward or downward movement within the scene, which is falsely attributed to compressions. This will not affect the live rate hugely, but it will affect the compression statistics on the summary page. False negatives (missed compressions) occur in two test videos. One reason for this is a large amount of lateral movement is mistakenly assumed to be the start of artificial ventilation, and subsequent compressions are missed before the algorithm self-corrects. The other reason is that compressions in one video did not meet the movement threshold.

The results for recognising breaths are also very good. Most notable is the fact that there are no missed breaths. If missed breaths were to occur it is possible that the overall rate would become inaccurate due to a large gap between compressions without recognising breaths in between them.

4.2.2 Evaluation of Chest Compression Rate and Chest Compression Fraction

In all cases, the chest compression rate is sampled every 2 seconds from the beginning of compressions in the ground truth. This means that the results will indicate the rate at the exact same points in time to give a fair comparison. The chosen interval is 2 seconds as this is the quickest option available for audio feedback on compression rates. The metrics used were both the mean compression rate calculated throughout each test video, but also the mean difference between each calculated rate and the percentage of rates within 3 compressions per minute of each other. It is not only important for the mean rates to be close, but for the majority of rates at any point in time to be close to one another. The results were very good (See Table 2), with the mean rates all within 1 compression per minute of each other, and over 95% of rates in each video being within 3 compressions per minute of the real value. This suggests the calculated rate is good enough to be considered an accurate assessment.

Discrepancies in the CCF between the vision algorithm and the ground truth are mainly due to false positives at the start and end of the CPR session. These false positives will falsely indicate that the CPR session was longer than it actually was, giving a greater overall fraction. The results for the CCF are encouraging, with most results being within 1% of the true result. In all cases, the computed CCF gives a reasonable idea of whether the compressions are being done for a suitable portion of time.

4.3 Comment on Results

The system gives extremely good results. Given that there are very few false positives or false negatives for compressions and breaths, it follows that the calculated CCR and CCF are also very close to each other. The vision algorithm is not only able to recognise breaths and compressions, but also of giving good results of the current rate at most points in time during the CPR administration.

5 CONCLUSIONS

The prototype system presented here successfully monitors CCR and breaths for people giving CPR. It
Table 1: True positives (TP), false positives (FP) and false negatives (FN) of compressions and breaths for each test video, along with their totals. In addition the overall accuracy, precision and recall for compressions and breaths are presented.

<table>
<thead>
<tr>
<th>Video #</th>
<th>Compressions</th>
<th>Breaths</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>1</td>
<td>144</td>
<td>0</td>
</tr>
<tr>
<td>2.1</td>
<td>127</td>
<td>0</td>
</tr>
<tr>
<td>2.2</td>
<td>174</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>169</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>186</td>
<td>1</td>
</tr>
<tr>
<td>5.1</td>
<td>211</td>
<td>3</td>
</tr>
<tr>
<td>5.2</td>
<td>211</td>
<td>0</td>
</tr>
<tr>
<td>6.1</td>
<td>175</td>
<td>1</td>
</tr>
<tr>
<td>6.2</td>
<td>200</td>
<td>4</td>
</tr>
<tr>
<td>6.3</td>
<td>170</td>
<td>1</td>
</tr>
<tr>
<td>6.4</td>
<td>201</td>
<td>2</td>
</tr>
<tr>
<td>Totals</td>
<td>1974</td>
<td>14</td>
</tr>
<tr>
<td>Overall</td>
<td>Accuracy</td>
<td>Precision</td>
</tr>
<tr>
<td></td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table 2: Mean CCR calculated from the ground truth and vision algorithm for each video, along with the proportion of rates which were with 3 compressions per minute. In addition the CCF from the ground truth and from the system are presented.

<table>
<thead>
<tr>
<th>Video #</th>
<th>Ground Truth Mean CCR</th>
<th>Computed Mean CCR</th>
<th>Proportion within 3 compressions/min</th>
<th>Ground Truth CCF</th>
<th>Computed CCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>108.00</td>
<td>108.15</td>
<td>1.00</td>
<td>0.773</td>
<td>0.773</td>
</tr>
<tr>
<td>2.1</td>
<td>103.80</td>
<td>103.90</td>
<td>0.95</td>
<td>0.958</td>
<td>0.958</td>
</tr>
<tr>
<td>2.2</td>
<td>73.95</td>
<td>73.60</td>
<td>0.93</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>3</td>
<td>123.27</td>
<td>122.98</td>
<td>0.99</td>
<td>0.558</td>
<td>0.555</td>
</tr>
<tr>
<td>4</td>
<td>125.55</td>
<td>125.74</td>
<td>1.00</td>
<td>0.700</td>
<td>0.700</td>
</tr>
<tr>
<td>5.1</td>
<td>115.10</td>
<td>114.28</td>
<td>0.93</td>
<td>0.696</td>
<td>0.701</td>
</tr>
<tr>
<td>5.2</td>
<td>114.98</td>
<td>114.60</td>
<td>0.96</td>
<td>0.731</td>
<td>0.732</td>
</tr>
<tr>
<td>6.1</td>
<td>92.50</td>
<td>92.50</td>
<td>1.00</td>
<td>0.800</td>
<td>0.796</td>
</tr>
<tr>
<td>6.2</td>
<td>92.17</td>
<td>91.60</td>
<td>0.95</td>
<td>0.795</td>
<td>0.807</td>
</tr>
<tr>
<td>6.3</td>
<td>92.84</td>
<td>92.69</td>
<td>1.00</td>
<td>0.800</td>
<td>0.791</td>
</tr>
<tr>
<td>6.4</td>
<td>92.22</td>
<td>91.59</td>
<td>0.96</td>
<td>0.771</td>
<td>0.807</td>
</tr>
</tbody>
</table>

also provides an accurate value of the CCF. As this is the first stage in a larger development it has not yet been tested for robustness with respect to the position of the camera relative to the person giving CPR.

The concept clearly has potential for use in training environments, where volunteers typically practice CPR every time they meet but rarely with precise feedback on CCR or CCF. In the longer term this technique might have potential for use in the field although a number of other issues need to be addressed (such as CCD, robustness, camera location requirements, etc.).

REFERENCES


