Person-independent discomfort detection system for infants

Citation for published version (APA):

Document status and date:
Published: 01/05/2016

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Person-Independent Discomfort Detection System for Infants

C. Li$^1$, S. Zinger$^1$, W. E. Tjon a Ten$^2$, P. H. N. de With$^1$

$^1$Eindhoven University of Technology, 5600MB, Eindhoven, the Netherlands
$^2$Maxima Medical Center, 5500MB, Veldhoven, the Netherlands

{c.li2, s.zinger, p.h.n.de.with}@tue.nl, w.tjonaten@mmc.nl

Abstract

Automatic discomfort detection for infants is important in healthcare, since infants have no ability to express their discomfort. We propose a video analysis system, based on supervised learning and classifying previously unseen infants from the testing set in a fully automated way. The first stage of our system consists of face-based face detection, and then fit a face shape to the detected face area by using a Constrained Local Model (CLM). In the second stage, we analyze expression features by using Elongated Local Binary Patterns (ELBP), and classify expression features with a Support Vector Machine (SVM) for discomfort detection. The key contribution of our system is that the face model is infant-independent by employing a Constrained Local Model without prior knowledge about previously unseen infants. The system detects discomfort with an accuracy of 84.3%, a sensitivity of 82.4%, and specificity of 84.9% on the testing set containing videos of 11 infants. In addition, in order to increase the robustness of the system to head rotation, we introduce a face recovery method based on the symmetry of the face. With this step, the previous performance parameters increase by 3.1 – 3.8% tested with videos of 2 infants containing 2,010 frames.

1 Introduction

Discomfort and pain assessment for infants is an important and challenging task, since infants are not able to verbally communicate their discomfort. Neglecting discomfort or pain of infants can cause problems in their further development [1]. Therefore, nurses or parents normally take the responsibility of monitoring and assessing pain of infants when they are admitted to a hospital. Several behavioral and physiological changes can indicate potential discomfort of infants, such as crying, facial tension, frequent body movements, heart rate, respiratory response, blood pressure, and levels of oxygen and carbon dioxide in the blood.

To assess pain, various infant pain scoring methods have been developed, such as the PIPP (Premature Infant Pain Profile) and the COMFORT pain scale, which are based on the above-mentioned clinical parameters. All nurses responsible for infant pain assessment have to be trained to apply a pain scoring, while observing infants for a certain time period for pain estimation. However, monitoring by nurses is time consuming, expensive and, as assessments are done in intervals, changes in pain/discomfort intensity between the intervals are not detected. Therefore an automated discomfort detection system is highly attractive. Besides monitoring infants for a long period, such a system opens new possibilities for reliable disease diagnosis, such as Gastro-Esophageal Reflux Disease (GERD).

Automatic discomfort detection for infants based on video analysis is a challenging task for several reasons. At first, infants are often sleeping, so that their eyeballs are invisible. Therefore, many face detectors based on finding eyes cannot be used for this task. Secondly, most infants need pacifiers to stay calm, and this partially occludes
the features presenting facial expressions. Another complication is that discomfort can 
be detected in specific cases only and may be tuned to a particular child, which harms 
the general use of the system [2]. Since facial expression is a major feature that helps 
professionals to assess pain of infants, much attention has been given to analyze facial 
expressions. In [3], the authors propose a facial expression recognition system using 
an automated eye detection for face detection, together with a combination of Gabor 
features and Support Vector Machine (SVM) for expression classification based on the 
face coding system (FACS). However, infants, especially neonates, lie in their bed with 
their eyes closed, and an automated eye detection will fail for them. Lucey et al. [4] 
proposed a system for a pain intensity estimation based on an Active Appearance Model 
(AAM). They first extract shape and appearance features from AAM, and then train 
separate SVM classifiers for action units with these features. However, this approach 
requires manual labeling of the key frames of a sequence, so that the system is not fully 
automated. In this paper, we present our video analysis system that automatically 
monitors discomfort, but now in a generic way, independent of the observed infant. 

We contribute to this research in two aspects. At first, our system is infant-
independent and requires no prior knowledge about the previously unseen infants. 
Second, our system is validated on a clinical dataset. The paper is organized as follows. 
Section 2 explains each component of the system in detail. In Section 3, experimental 
results are shown for the system evaluation. Finally, conclusions are drawn in Section 4.

2 System Design

The system consists of two main components: face detection and discomfort detection, 
as shown in Fig. 1. The first part employs a combination of the Viola-Jones face de-
tector and a skin-color detector for finding the face area. After that, we describe the 
shape of the face with a Constrained Local Model (CLM). Based on the concept of 
similarity-normalized appearance (SAPP), Elongated Local Binary Patterns (ELBP) 
are extracted as facial expression features. For classification, an Support Vector Ma-
chine (SVM) is used for distinguishing discomfort from comfort. A more detailed 
description of each block now follows.
2.1 Face Detection

In order to optimize the fitting of a shape model to the image, we first use a combination of a Viola-Jones face detector and a Gaussian mixture model skin color detector for locating the face area as proposed in [5]. In our system, we analyze frames in the YCbCr color space [2]. A false detection of a face directly causes a false detection of discomfort. Therefore, the system selects the detected area in the following way. If the detected area of a frame is smaller than a pre-defined threshold, then the system regards this detection as a false detection and discards this frame. Only the frames with the detected area larger than the threshold are passed on to the next stage. Then within the detected face area, we apply CLM for aligning a face shape to the frame.

A Constrained Local Model is a person-independent model compared to AAM [6]. It is a joint shape and texture model introduced by Cristinacce et al. in [7]. The shape is represented as a concatenated vector of $X$ and $Y$ coordinates, as follows:

$$X = (X_1, ..., X_n, Y_1, ..., Y_n)^T, \quad (1)$$

where $n$ is the number of points in the shape of the CLM. The shape $X$ can be expressed as a base shape $\bar{x}$ and a linear combination of shape vectors $\Phi$, denoted by

$$X = T_t(\bar{x} + \Phi b), \quad (2)$$

where the coefficients $b = (b_1, ..., b_m)^T$ are the shape parameters. The shape model is normally computed from training data consisting of a set of images with the shape points marked manually. To obtain the base shape $\bar{x}$ and the shape variation $b$, the Procrustes alignment algorithm and Principle Component Analysis (PCA) are commonly used and also adopted here. To align the orientation and face scale, transformation $T_t$ is introduced to cover these properties and it represents a similarity transform with the parameter $t$.

The CLM application is novel to the discomfort detection for infants with the described challenging conditions (pacifiers, no eyeballs). Hence, the fitting of the model is crucial. To fit the shape model to an image, a response map of each patch model is obtained by an exhaustive local search for each landmark around an initial shape using a feature detector. Then the shape parameters $b$ are optimized by criteria that jointly maximize the detection responses over all the landmarks. In our system, we adopt the optimization strategy based on subspace-constrained mean-shifts, proposed by Saragih et al. [8]. An example of an original frame and a corresponding aligned face shape by CLM are shown in Fig. 2(a) and Fig. 2(b).

2.2 Discomfort Detection

Discomfort detection is performed in two steps: feature extraction and feature classification. These steps are described below.

Feature Extraction

Once the shape model is aligned to the image by CLM, facial features can be obtained based on similarity-normalized shape and SAPP defined by Lucey et al. [4]. SAPP refers to the face appearance after the removal of the rigid geometric variations and scale. This is achieved by warping the pixels of characteristic regions of the facial image into the similarity-normalized shape. A corresponding SAPP obtained from the face shape of Fig. 2(b) is displayed in Fig. 3(a). However, if the face is not frontal, the CLM fitting for the occluded face is not sufficiently accurate (see e.g. Fig. 2(c) and Fig. 2(d)). As a consequence, a distorted SAPP is obtained by the misalignment of the face shape (Fig. 3(b)). For the automated detection system, our contribution is to improve the robustness in applying the CLM when the faces are partly occluded. The
Figure 2. (a) Frame of the original video with frontal face; (b) Detected face shape of (a); (c) Frame of the original video with head rotation; (d) Detected face shape of (c);

Figure 3. (a): SAPP of the image shown in Fig. 2(a). (b): Distorted SAPP caused by face shape misalignment of the image shown in Fig. 2(d). (c): Restored SAPP of (b) appearance in Fig. 3(b) is containing distortions at the wrinkles at the right eye and right side just above the mouth. In order to remove such errors, we create a frontal face orientation, and mirror the pixels from its visible part to the area where the face is occluded, thereby exploiting the symmetry of the face. This leads to a better restored SAPP, of which an example of is shown in Fig. 3(c).

ELBP is shown to have a very good performance on face recognition [9] [2]. Therefore, we extract ELBP features on SAPP for discomfort detection in our system. For ELBP calculation, the neighborhood pixels of the center pixel are defined as an ellipse with minor radius 1 in the vertical direction and major radius 2 in the horizontal direction around that pixel. The ELBP is calculated by comparing the gray value of the center pixel with the surrounding neighborhood pixels, as specified below:

\[
ELBP(x) = \sum_{i=1}^{P} d(g_i - g_c)2^{i-1}. \tag{3}
\]
Here parameter $P$ denotes the number of the neighborhood pixels and $g_c, g_i$ denote the gray values of the center pixel and the neighborhood pixels, respectively. The binary output function $d$ is defined by

$$d(g_i - g_c) = \begin{cases} 
1, & \text{for } (g_i - g_c) \geq 0; \\
0, & \text{for } (g_i - g_c) < 0.
\end{cases}$$  
(4)

**Feature Classification**

SVM has been demonstrated to be useful in a number of facial recognition and expression recognition tasks [10]. This classifier defines the optimal hyperplane by maximizing the margin between positive and negative samples for a specified class. For detecting discomfort, we train an SVM with a linear kernel to distinguish discomfort from comfort, based on ELBP features extracted from SAPP, as described above, with the LIBSVM library [11]. The combination of the previous algorithm steps enables automated discomfort detection.

### 3 Experiments

In this section, we first introduce the databases used for training and testing our system. Then we describe the applied evaluation metrics. Finally, we present key performance parameters: accuracy, specificity and sensitivity for face detection and discomfort detection in our system.

#### 3.1 Database

For training the discomfort classifier, we re-use the database from Fotiadou et al. [2]. This database consists of 15 videos of 8 infants, 5 displaying comfort, 1 discomfort and 9 videos containing both. Each video has a frame rate of 25 frames per second with a resolution of 1920 $\times$ 1080 pixels, which is denoted as Database I. For the purpose of testing, we also use a database of 106 videos from 38 infants that have been recorded at the Maxima Medical Center (MMC), Veldhoven, the Netherlands, by Slaats et al. [5]. Videos of the faces of infants experiencing pain resulting from various interventions, like a heel prick, placing an intravenous line, a venipuncture, a vaccination or from post-operative pain, are recorded. However, this database includes various children and situations, such as premature neonates, infants with tubes and pacifiers, and occlusions on the face. Therefore, we choose a collection of 13 videos from 11 infants without occlusion of the face, which we call Database II. From the 13 videos, 5 videos display discomfort, 7 videos display comfort and 1 video displays both. Each video has a frame rate of 30 frames per second and a spatial resolution of 1280 $\times$ 720 pixels. A description of both datasets is shown in Table 1. Examples of video frames from both databases are portrayed by Fig. 4.
Table 1. Database descriptions

<table>
<thead>
<tr>
<th></th>
<th>Infants</th>
<th>Videos</th>
<th>Total Frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database I</td>
<td>8</td>
<td>15</td>
<td>43,823</td>
</tr>
<tr>
<td>Database II</td>
<td>11</td>
<td>13</td>
<td>13,917</td>
</tr>
</tbody>
</table>

Figure 5. (a) Manually labeled ground-truth image; (b) Detected face surrounded by a rectangle.

3.2 Evaluation Results

We evaluate our system in two aspects: face detection rate and discomfort detection rate. For face detection, we consider a detection to be correct when a rectangle encompassing the face mesh fitted in the image has a 70% overlap area with the rectangle in the ground-truth image. A ground-truth image has a manually annotated rectangle encompassing the eyes, nose and mouth. Fig. 5 shows an example of a manually labeled ground-truth image and a face mesh surrounded by a rectangle.

For discomfort detection, we evaluate the SVM classifier based on single frames. For each video frame in the database, the ground-truth of discomfort is provided by nurses from the Veldhoven MMC, who are experienced in giving pain scores. The performance of discomfort detection is obtained by comparing the output of the SVM classifier with the ground-truth. To measure the performance, we calculate accuracy, specificity and sensitivity of the classification results.

3.2.1 Face Detection

The Viola-Jones detector is trained with 25 infants from the database obtained by Slaats et al. [5], excluding the infant videos we use from Database II. Table 2 shows the performance of our CLM-based face detection system evaluated with Database I and Database II. In this table, “total frame” means all frames of each dataset, and “output frames” means frames with a face mesh detected by the face detection stage. It shows that the true detection rate over all frames is 70.9% and 81.5%. However, more importantly, the rate of true detection over the output from the CLM-based face detector is 91.6% and 96.9% for Database I and Database II, respectively.

3.2.2 Discomfort Detection

We have used infants included in Database I proposed by Fotiadou et al. [2] for training the discomfort detection classifier. For testing the performance of the discomfort detection, we apply infant videos from Database II. These infants are all unseen subjects to our discomfort classifier. Table 3 shows the performance of ELBP features extracted from SAPP for discomfort detection evaluated with Database II without face interpolation. It can be observed that our system achieves an accuracy of 84.3%, with a
Table 2. Performance of CLM-based face detection evaluated with Database I and Database II.

<table>
<thead>
<tr>
<th></th>
<th>Database I</th>
<th>Database II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Output Frames/Total Frames</td>
<td>77.1%</td>
<td>84.3%</td>
</tr>
<tr>
<td>True Detection/Total Frames</td>
<td>70.9%</td>
<td>81.5%</td>
</tr>
<tr>
<td>True Detection/Output Frames</td>
<td>91.9%</td>
<td>96.9%</td>
</tr>
</tbody>
</table>

Table 3. Performance of CLM-based ELBP+SAPP features evaluated with Database II.

<table>
<thead>
<tr>
<th>Database II</th>
<th>Accuracy</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>84.3%</td>
<td>82.4%</td>
<td>84.9%</td>
</tr>
</tbody>
</table>

Table 4. Comparison of ELBP+original SAPP and ELBP+mirrored SAPP.

<table>
<thead>
<tr>
<th></th>
<th>3 videos of 2 infants from Database II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ELBP+original SAPP</td>
</tr>
<tr>
<td>Accuracy</td>
<td>73.8%</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>85.6%</td>
</tr>
<tr>
<td>Specificity</td>
<td>67.7%</td>
</tr>
</tbody>
</table>

sensitivity of 82.4% and a specificity of 84.9% for unseen infants for discomfort detection. In order to evaluate the impact of face interpolation, we have manually chosen 3 videos of 2 head-rotated infants from Database II with a total of 2,010 frames, containing significant head rotations and movements. These videos form a worst-case test. This explains the lower performance scores in Table 4, which also shows that, with face interpolation, the accuracy increases by 3.6%, sensitivity by 3.8%, and specificity by 3.1%.

4 Conclusion

In this paper, we have proposed an automated and person-independent system to detect discomfort for infants. The proposed algorithm exploits CLM for infant face detection in a robust way without any prior knowledge. The robustness improvement is achieved by (1) training the CLM with a generic model that is person-independent, and (2) by solving misalignment errors in the model via face mirroring. The improved robustness brings the practical use of the system in a hospital environment where infants come and leave continuously, closer to reality. The system can detect discomfort with a score of an accuracy of 84.3%, a sensitivity of 82.4%, and a specificity of 84.9% for Database II, however, these numbers were not yet obtained for the full algorithm. False positives and false negatives for discomfort classification are mainly due to the misalignment of the fitting face shape to the image. With face interpolation, the system is more robust to head orientation. In near future, we will experiment on larger datasets, and extend the system to gastro-esophageal reflux disease patients, in order to assist the diagnosis procedure with our classification results.
References


