ULTRASOUND VIDEO ANALYSIS FOR UNDERSTANDING INFANT BREASTFEEDING

Gianluca Monaci
Video & Image Processing, Philips Research
Eindhoven, The Netherlands

Mike Woolridge
School of Healthcare, University of Leeds
Leeds, UK

ABSTRACT

While it has been widely proved that breastfeeding is the healthiest feeding option for a baby and its mother, the mechanisms by which a baby removes milk from the breast are still not completely known. Partly this is due to the lack of tools to analyze images of the infant oral cavity during feeding automatically and quantitatively. In this paper we propose two methods for analyzing ultrasound videos to automatically detect relevant events such as sucking and swallowing of milk and to discriminate different types of tongue action during milk removal. The proposed algorithms provide, for the first time, quantitative indications of the type of activities carried out during breastfeeding by the baby, promising unprecedented advancements in the field.

Index Terms— Ultrasonic image processing, Image sequence analysis

1. INTRODUCTION

In this paper we introduce two novel methods to automatically detect and classify relevant events in ultrasound videos of breastfeeding infants. It has been widely demonstrated that breastfeeding is healthier than formula or bottle feeding, both for a baby and its mother. Studies have shown that breastfed infants have fewer ear, respiratory, gastrointestinal infections [1]. Breastfeeding has positive effects on the development of the infant’s oral cavity [2], while breastfeeding mothers are less likely to develop breast cancer [3]. Unfortunately, they often stop due to problems they encounter, for example during the establishment of breastfeeding or later when they want to combine breastfeeding with working [4]. A better knowledge of breastfeeding mechanisms will lead to better alternatives to help these mothers prolong breastfeeding.

There is an ongoing debate about the mechanisms used by breastfeeding infants to remove milk from the breast. It has long been agreed that infants predominantly extract milk by drawing the nipple into the mouth and performing a cyclic wave-like pressure on the nipple, known as peristaltic action [5]. In a recent study [6], the authors challenge this view asserting that infant’s tongue does not exhibit a peristaltic motion, but moves up and down, creating vacuum in the baby’s palate in front of the nipple, inducing milk flow. In existing studies only subjective, qualitative observations of the infant’s tongue motion have been reported, making it difficult to objectively assess the relevance of peristaltic and vacuum actions to milk removal. Automatic tools to analyze the tongue motion and provide quantitative measures of these phenomena are required to advance the knowledge of breastfeeding mechanisms.

In this work we describe methods devised to analyze video sequences of ultrasound sub-mental scans of the midline of the infant’s oral cavity during breastfeeding, recorded at the Leeds University School of Healthcare. Ultrasound imaging is widely used to analyze infants’ breastfeeding because it is safe and non-invasive [6, 7]. All the results shown are based on pilot analyses of an archetypal ultrasound image sequence, selected because it exhibited the diversity of feeding styles typically shown on the breast. This sequence was subjectively labelled by a specialist in this area (MW), and consisted of about 1500 consecutive ultrasound images. One frame from this ultrasound video is shown in Fig. 1 (left). The main anatomic structures of the infant’s oral cavity are roughly defined and labelled in Fig. 1 (right). In the recorded sequences, the infant is typically performing one of these three activities: i) Doing nothing, ii) Stimulating the mother’s nipple, without swallowing milk (non-nutritive sucking or simply sucking), iii) Sucking and swallowing milk (nutritive sucking, or swallowing). The major challenge in this study
is to automatically detect and characterize these activities in what are inherently noisy ultrasound images, such as that shown in Fig. 1.

This paper tackles these challenges utilising two main contributions:

1. In Sec. 2 we present an algorithm to automatically detect and classify sucking and swallowing activities. This will help quantifying infants’ activities during breastfeeding. Besides, the proposed method will be useful for all sorts of studies in the field, as it makes more efficient the otherwise time-consuming and tedious visual inspection of ultrasound video material.

2. In Sec. 3 we introduce a novel method to discriminate between peristaltic and vacuum action. This second problem is far more challenging, as the distinction between peristaltic and vacuum action is still debated between pediatricians and experts in the field. We will show that the results of proposed approach are extremely promising.

2. DETECTION OF NUTRITIVE AND NON-NUTRITIVE SUCKING

The first observation we made is that the overall intensity of the ultrasound images varies through time in a way that is closely related to the infant’s sucking rhythm. Figure 2 (top) shows the variation over time of the average intensity of the whole ultrasound frame for a part of the test sequence in Fig. 1. The most marked local minima in this raw signal match closely periods of infant activity in the video, such as non-nutritive and nutritive sucking.

Low frequency variations of this intensity signal are mainly due to global changes in the image and reflect the motion of the whole infant’s head. High-frequency oscillations instead are mainly due to noise. We thus filter the average intensity signal with a bandpass, fourth order, Butterworth filter with band between $0.08 \cdot f_N$ and $0.24 \cdot f_N$, where $f_N$ is the Nyquist frequency. Here the video is recorded at 25fps, thus $f_N = 12.5\text{Hz}$. The average intensity signal filtered with such a filter, let us call it $s$, is plotted in Fig. 2 (bottom). In the figure, local minima smaller than a threshold $T = -0.8 \cdot \text{std}(s) \approx -1$ are also marked with red squares ($\text{std}(\cdot)$ denotes the standard deviation of the signal). These points closely correspond to points in the video where the infant is performing a sucking or swallowing action, although it is not possible to discriminate between these two actions analyzing only this global feature.

To provide a more accurate analysis of the subtle phenomena we want to model, we developed a similar approach where image intensity variations in different regions of the scan are analyzed. The majority of the sucking activity is in the upper half of the scan, and in this analysis the goal is to demarcate discrete areas where specific events are taking place. Three areas of interest over which the analysis is focused are defined on the image in red (left), green (middle) and blue (right) in Fig. 3. The red sector comprises the nipple and the frontal part of the tongue. The green sector discretely focuses on the area just beyond the tip of the nipple, and comprises the posterior part of the tongue and the soft palate. The blue sector is a broad, non-specific area, where we attempt to pick up swallowing in a fairly generic manner.

In the implemented application, the three sectors are selected by hand for each sequence and kept fixed in size and position. While these areas are fixed, there is typically a lot of motion due to movement of the infant. If, due to this motion, a specific anatomical structure bridges the junction between two adjacent sectors, then its movement would contribute to both sectors simultaneously, contaminating the resulting analysis. We try to avoid that by creating a gap between the three areas.

The average image intensity is computed for each region

---

**Fig. 2.** Average ultrasound image intensity plotted as a function of time (top) and its bandpass filtered version (bottom). Local minima are marked with red squares.

**Fig. 3.** The three regions over which the analysis is focused. The nipple and the frontal part of the tongue are on the red region (left). The posterior part of the tongue and the soft palate are in the green area (middle). The blue region (right) is a broad area where we want to pick up swallowing.
and the signals are filtered with the bandpass filter described above. The raw and filtered average intensity signals for the three areas highlighted in Fig. 3 are shown in Fig. 4. The signals are captured over the same time interval of Fig. 2.

Both raw data and filtered signals are similar to those representing the entire image intensity variations (Fig. 2). However, this region-based analysis allows to make few interesting observations. First, the signals from the red and green areas evolve in a very similar way, and local minima in these signals reflect closely infant’s sucking activity. Along the whole sequence the average signal from the red region seems cleaner (with less spurious peaks) than the green one, representing better the actual baby activity. We detect non-nutritive sucking events by detecting local minima in the bandpass filtered intensity signal from the red area. We indicate this signal with \( r \). To filter out peaks due to noise, only local minima smaller than the threshold \( T_1 = -0.8 \cdot \text{std}(r) \approx -0.8 \) are considered.

The signal from the blue region clearly picks up the swallowing activity. The intensity of this signal changes through time in a way that is similar to the green and red signal, with one clear exception, visible in Fig. 4. When a nutritive sucking event (swallowing) occurs, the intensity signal of the blue region drops drastically, as in the raw signal in Fig. 4 (left column, bottom row), around frame 460. The swallowing activity is detected by detecting particularly low local minima in the filtered average intensity signal from the blue area, indicated with \( b \). Only local minima smaller than the threshold \( T_2 = -1.5 \cdot \text{std}(b) \approx -2 \) are detected.

A specialist in this area (MW, School of Healthcare, University of Leeds) labelled non-nutritive and nutritive sucking events in the test sequence. In the first part of the sequence the baby stimulates the nipple by repeatedly sucking it. In this part however, very few swallows are present. After, the infant enters a period during which each suck corresponds to a swallow. In total there are 71 sucks and 32 swallows. Using the automatic technique described above, 71 sucks and 25 swallows are detected. Of the 71 sucks detected, 65 are correctly detected, while 6 are missed (false negative) and 6 sucks are positively detected while no sucking was performed (false positive). The false negative detections occur because the local minima are bigger than the threshold \( T_1 \), while the false positive detections occur in the initial part of the sequence where the baby is moving while the scan is initiated. Of the 32 swallows present in the sequence, 25 are correctly detected and 7 are missed (false negative) because the local minima in the blue signal is larger than the threshold \( T_2 \). Five of these errors occur in the last 150 frames of the sequence, where the infant in the scan starts to move and considerable motion noise is present in the sequence. No false positives occur in this case.

Considering that nutritive and non-nutritive sucking events are difficult to detect manually and classify even for experts, the detection results obtained on the test sequence are extremely promising. The critical parameters here are the peak detection thresholds which, for simplicity, are linked to the standard deviation of the signals. The use of adaptive thresholding methods to improve the detection results, e.g. [8], will be the object of future investigation.

### 3. CLASSIFICATION OF PERISTALTIC AND VACUUM ACTIONS

In the test sequence we are considering, in the first phase the infant sucks the nipple to stimulate milk release. In this phase, it appears clear that the baby’s tongue moves following a characteristic peristaltic motion, as described in [5]. After this the infant enters the phase in which a swallow corresponds to each suck. During this phase there is a more pronounced vacuum action and the tongue assumes a distinctly different shape, as described in [6]. Interestingly, the average intensity traces from the red and green regions change substantially when the infant goes from one regime to the other. To distinguish between peristaltic and vacuum action we propose to analyze the phase-shift between the filtered average

---

**Fig. 4.** Temporal evolution of the average intensity (left column) and filtered average intensity (right column) of the red (1st row), green (2nd row) and blue (3rd row) regions as highlighted in Fig. 3.
Fig. 5. Phase shift $M$ between the filtered average intensity signals from the red and green regions of Fig. 3, plotted as a function of time. During peristaltic action, $M \leq 0$, while during vacuum action $M > 0$.

intensity signals from the red and green regions. Let us call them $r$ and $g$ respectively. The phase-shift between the filtered intensity signals is estimated by computing the cross-correlation between the signals using different time lags. At each time instant $t$, $N = 200$ samples of each signal (or as many are available if $t < 200$) are buffered. For each time lag $m \in [-10, 10]$, the cross-correlation between the signals $r$ and $g$ of length $N$ can be written as:

$$R_{rg}(m) = \sum_{n=0}^{N-m-1} r(n)g(n + m). \quad (1)$$

The phase shift $M$ between $r$ and $g$ is then estimated as the time shift at which the correlation between the two waveforms is maximum:

$$M = \arg \max_{m \in [-10, 10]} R_{rg}(m).$$

Figure 5 shows the estimated phase-shift $M$ as a function of time for the test sequence. One author (MW) annotated the sequence as shown in Fig. 5: during the first 1000 frames, the dominant action performed by the infant is peristaltic, while in the remaining 500 frames of the video the dominant action is closer to 'vacuum' generation. During this first phase, intensity peaks in the red region precede or are in phase with those in the green region, as shown by the phase-shift $M$ between $r$ and $g$ being negative or null. When going into the vacuum action regime, related to more intense swallowing activity, intensity peaks in the green region typically precede peaks in the red zone and $M > 0$. The phase-shift between the two signals seems to reflect the action exercised by the baby during breastfeeding. This phase switch is probably due to the change in configuration of the tongue during the two phases: during peristaltic action the tongue profile designs a wave travelling from left to right [5]. During vacuum action the infant’s tongue essentially moves up and down with a flatter profile [6].

The proposed technique seems capable, for the first time, to quantify the relative amount of peristaltic and vacuum action during feeding by simply measuring the phase shift between the $r$ and $g$ signals. These observations are currently being validated on a large set of ultrasound video scans.

4. DISCUSSION

In this paper we have presented two ultrasound video analysis techniques developed to help shed some light on the mechanisms underlying infant’s breastfeeding. We first introduced a new algorithm to automatically detect swallowing and sucking events through the analysis of image intensity in selected regions of interest. We also proposed a methodology to discriminate peristaltic and vacuum tongue action based on the analysis of the phase-shift between the average image intensity signal emerging from the frontal and the posterior part of the infant’s tongue. Results on a labelled test sequence of more than 1500 ultrasound images are extremely encouraging: the proposed methods provide, for the first time, quantitative indications of the type of activities carried out during breastfeeding by the infant. Given the number of images analyzed here, current and future work is focused on validating the findings presented in this paper by applying the proposed methods on a larger dataset of ultrasound video sequences collected at the Leeds University School of Healthcare.

5. REFERENCES