Extraction of Periodic Features from Video Signals

Davide Alinovi and Riccardo Raheli

Department of Engineering and Architecture
Information Engineering Unit
University of Parma
Email: {davide.alinovi, riccardo.raheli}@unipr.it

***Abstract*—In a number of application scenarios, proper video signals may exhibit simultaneous correlation characteristics over the space and time dimensions which jointly describe periodic features or behaviors. Examples of such scenarios may be found in video monitoring of physical systems, sport and athlete coaching with automatic video supervision, biomedical applications to newborn video monitoring for the detection of epileptic seizures or apnea episodes, surveillance systems and others. A general Maximum Likelihood (ML) approach to the detection of common periodic features possibly present in a set of video signals and the estimation of their characteristics, such as the fundamental frequency and the local amplitude, is proposed. Application examples in various scenarios are presented and the performance of the proposed ML solutions is shown to be effective.**

Keywords–Features extraction; periodicity analysis; video processing; maximum likelihood estimation.

I. INTRODUCTION

A video signal is characterized by a multidimensional domain in which two space dimensions specify the pixel position within a frame and a time dimension describes the evolution of the frame image (3D). An additional space dimension may come from the simultaneous use of multiple cameras framing the same scene from different viewpoints, bringing the overall dimensionality to 4D. This paper discusses the extraction of periodic features from video signals obtained by one or multiple cameras—a topic which has been the subject of a large body of literature, e.g., see [1] and references therein.

A first approach considered in the literature uses spatial matching to identify an object or a portion of the image, follow the evolution of its trajectory over time in successive frames and analyze this trajectory to extract possible periodic features [1], [2]. Despite being very general, this approach is impacted by the reliability of the spatial matching step, which is largely affected by the quality and resolution of the video sequence, as well as possible optical effects, including illumination variations, reflections, occlusions and others.

As nicely pointed in [2, Figure 1], the fundamental ability to recognize periodic features in a sequence of frames does not require high quality or resolution, as demonstrated by an example of a significantly blurred low-resolution sequence of images in which the human brain can still appreciate the periodic feature of a man waking on a treadmill.

A second approach discussed in the literature avoids the critical spatial matching step and uses suitable projections of the video sequence in the spatial domain to extract compact representations of the video variations in the time domain, which can then be easily analyzed in the frequency domain to recognize possible periodic features. In this category, [1] projects each frame onto the x and y dimensions to obtain two signals $x[n]$ and $y[n]$, in the time index n, that can be jointly analyzed to extract possible periodic components. Another example within this approach is [3], where each frame is projected onto the single space dimension represented by the average luminance signal, which can be easily processed along the time domain to extract the frequency components of interest and detect possible periodic features.

These approaches, despite being general and reasonable, are based on specific initial assumptions—spatial matching in the first one and frame projection in the second one—which may possibly limit their effectiveness and efficiency. To avoid these specific assumptions and their possible consequences, in this paper, we wish to take a more basic and radical approach by considering the direct application of fundamental estimation and detection criteria to the multidimensional video signal. To this purpose, we have selected the sound and trustable Maximum Likelihood (ML) principle, which is very well studied, documented and widely applied [4], [5].

As in all applications of the ML criterion, a key part of the problem is the selection of a suitable observation model. To this purpose, we propose to base the estimation and detection process on the direct observation of the 3D or 4D video sequence, possibly affected by noise. The proposed solutions are then considered in a few application examples and their performance is analyzed and discussed.

The remainder of this paper is organized as follows. In Section II, the model of periodic variations and the extraction of related features in multidimensional video signals are described. Section III presents some applications in specific fields of the proposed technique. Finally, conclusions are drawn in Section IV.

II. ANALYSIS OF PERIODIC VARIATIONS

A. Preliminaries

A digital video signal consists of a series of digital images, also known as frames, properly captured over time. Precisely, a video can be defined as a multidimensional signal which describes the evolution over time occurring in the framed area. Moreover, digital frames are bidimensional projections of the real world on the camera sensor and a loss of information about the full 3D motion has to be considered. Therefore, a physical 3D displacement in the space corresponds to spatio-temporal variations in the structure of the multidimensional signal; such movements affect to some degree the pixel intensity values in the video signal captured by the camera sensor.

Periodic variations can be of particular interest: in fact, they can represent specific events and need to be properly detected

 ISBN: 978-1-61208-548-7

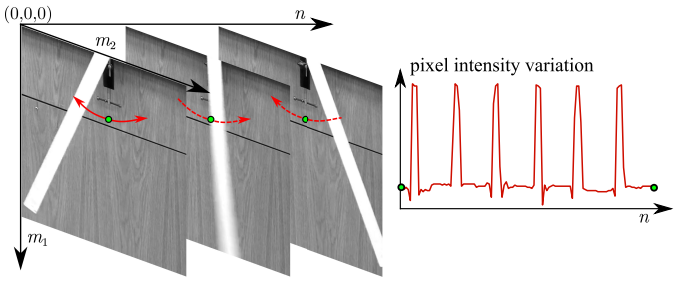


Figure 1. An example of periodic motion and the variations it causes in a multidimensional signal: a physical pendulum swinging from left to right.

and analyzed. In the specific case of periodic movements or recurring events, the pixel intensity values in the video stream may exhibit periodic features. As an example, in Figure 1, some frames extracted by a video capturing the motion of a physical pendulum are shown, with a spatio-temporal reference system describing the spatial plane (m_1, m_2) and the temporal dimension (n). Considering a reference pixel, specified by the highlighted point, the intensity variations which affect it are also shown: these periodic pixel-wise variations can be exploited to analyze spatio-temporal movements.

Before detailing the model of multidimensional periodicity and the approach to detection and estimation of periodic features, some preliminary considerations have to be introduced. The model proposed in the next subsection is valid under the following assumptions:

1) the camera is still or is moving solidly with the framed subject;
2) the subject of interest is not affected by translation or superposed motion.

The first condition is related to the considered scenario: as the main goal is to extract periodic features of subjects in the scene, global movements are not considered.[1] The second condition assumes that no large movements or intensity variations affect the subject: the only main motion components are expected to be the periodic variations.

B. Model of Periodic Variations

For the presented assumptions, the model of periodicity in a multidimensional signal is now defined. Consider a video signal acquired by a camera sensor with a sampling period T_s, namely with a frame rate $f_s = 1/T_s$. A gray-scale frame captured at the sampling instants nT_s can be described by a matrix $\mathbf{X}[n]$ composed of $M_1 \times M_2$ pixels, where M_1 and M_2 are the numbers of rows and columns of the matrix, respectively, and $X[m_1, m_2, n]$ is the intensity of the pixel with coordinates (m_1, m_2) in the n-th frame. For color videos, a proper number of channels has to be considered: as an example, for standard Red, Green and Blue (RGB) cameras each frame is composed by three matrices, one per color channel.

To simplify the notation for the following steps, the operator of vectorization of a matrix and its inverse are now introduced. Let $\mathbf{X}[n]$ be the matrix representing the video signal: the vectorized version $\mathbf{x}\,[n]$ is defined as

$$\begin{aligned}\mathbf{x}\,[n] = \quad & \mathrm{vec}\,(\mathbf{X}\,[n]) = \\ = \quad & \big[X[0,0,n] \;\cdots\; X[0,M_2-1,n] \\ & \;\; X[1,0,n] \;\cdots\; X[1,M_2-1,n] \\ & \qquad\qquad \vdots \\ & \;\; X[M_1-1,0,n] \;\cdots\; X[M_1-1,M_2-1,n]\big]^{\mathsf{T}}\end{aligned} \tag{1}$$

where the column vector $\mathbf{x}\,[n]$ has size $M_1 M_2 \times 1$, its element $x\,[p, n]$ denotes the intensity value of the p-th element of the n-th vectorized frame and $(\cdot)^{\mathsf{T}}$ denotes the vector transpose. Accordingly, the inverse operator is defined as

$$\mathbf{X}\,[n] = \mathrm{vec}^{-1}\,(\mathbf{x}\,[n]) \tag{2}$$

where the frame sampled at discrete time nT_s is retrieved to the original size $M_1 \times M_2$.

Another useful representation is given by the variations of the single p-th element over time. Starting from the vector $\mathbf{x}\,[n]$ introduced in (1), the evolution of the signal relative to the pixel in position p is denoted by the vector

$$\tilde{\mathbf{x}}\,[p] = \big[x[p,0]\; x[p,1] \;\ldots\; x[p,N-1]\big]^{\mathsf{T}} \tag{3}$$

which has size $N \times 1$, where N is the total number of considered frames.

Relying on the assumptions introduced at the beginning of Section II, the video frames are recorded by still cameras and contain pixel intensity variations related only to the periodic motion. In order to extract periodic features from the video signal, a proper model of the multidimensional structure is needed. Considering the scenario in which movements are driven by a single common periodicity, a useful model, including noise on the sequence of frames, may be given as

$$\mathbf{X}[n] = \mathbf{B} + \mathbf{A}\cos\left(2\pi f_0 n T_s + \mathbf{\Phi}\right) + \mathbf{W}[n] \tag{4}$$

where all the matrices have size $M_1 \times M_2$ (equal to the resolution of the involved camera sensor), $\mathbf{B}$ describes the pixel-wise continuous components, $\mathbf{A}$ is the matrix of the amplitudes, f_0 is the common fundamental frequency, T_s is the video sampling period, n is the frame index, $\mathbf{\Phi}$ is the matrix of the initial phases, $\{\mathbf{W}\,[n]\}$ are matrices of independent and identical distributed (i.i.d.) zero-mean Gaussian noise samples. In (4) and the following equations, the $\cos(\cdot)$ operator and the addition of a scalar to a vector or matrix are applied element-wise. The vectorized version of equation (4), according to (1), is given by

$$\mathbf{x}[n] = \mathbf{b} + \mathbf{a}\cos\left(2\pi f_0 n T_s + \boldsymbol{\phi}\right) + \mathbf{w}[n] \tag{5}$$

where the definition (1) is applied to the matrices $\mathbf{X}[n]$, $\mathbf{B}$, $\mathbf{A}$, $\mathbf{\Phi}$ and $\mathbf{W}[n]$.

Given this multidimensional model, the aim is to efficiently extract periodic features, such as the fundamental frequency f_0 and the amplitudes $\mathbf{A}$, which are useful to check the presence/absence of periodicity (or measure its repetition period) and identify the position of periodic variations in the video, respectively. In order to achieve the estimation of these parameters, it will be shown that the application of the ML approach to the model (5) is a reliable solution.

[1] Motion compensation algorithms could be used to limit this effect [1].

 ISBN: 978-1-61208-548-7

It can be noticed that extensions to a full RGB video, considered analyzing jointly the three color channels or multiple camera sensors, can be an application example of this approach, as shown in [6].

C. Generalized Maximum Likelihood Estimation

The approach consists of a generalized version of ML estimation applied to multidimensional signals. The parameters to be estimated are: the fundamental frequency f_0, the relative local amplitudes $\mathbf{a}$ and possibly the phases $\boldsymbol{\phi}$. These parameters can be collected in a vector $\boldsymbol{\theta} = [\mathbf{a}, f_0, \boldsymbol{\phi}]$. Following standard methods in [5], the likelihood function to be minimized in order to obtain the ML estimate $\hat{\boldsymbol{\theta}}$ is

$$J(\boldsymbol{\theta}) = \sum_{p=0}^{M_1M_2-1} \sum_{n=0}^{N-1} \Big[x[p,n] - a[p]\cos\left(2\pi f_0 n T_s + \phi[p]\right) \Big]^2 \tag{6}$$

where NT_s is a suitable observation window and $x[p,n]$ represents the observed video signal in the p-th position at discrete time nT_s.

The ML estimation of the parameters of interest is now derived, following proper steps similar to the ones in [5], [7]. Using trigonometric identities in (6), it is possible to obtain

$$J(\boldsymbol{\theta}) = \sum_{p=0}^{M_1M_2-1} \sum_{n=0}^{N-1} \Big[x[p,n] - \alpha[p]\cos\left(2\pi f_0 n T_s\right) - \beta[p]\sin\left(2\pi f_0 n T_s\right) \Big]^2 \tag{7}$$

where $\alpha[p] = a[p]\cos(\phi[p])$ and $\beta[p] = -a[p]\sin(\phi[p])$. As $a[p]$ and $\phi[p]$ are strictly related with $\alpha[p]$ and $\beta[p]$, it is possible to substitute the vector parameter $\boldsymbol{\theta}$ with $\boldsymbol{\theta}' = [\boldsymbol{\alpha}, \boldsymbol{\beta}, f_0]$. By properly combining the variables in the temporal dimension, it is possible to obtain a simplified version of the likelihood function:

$$J(\boldsymbol{\theta}') = \sum_{p=0}^{M_1M_2-1} (\tilde{\mathbf{x}}[p] - \alpha[p]\mathbf{c} - \beta[p]\mathbf{s})^{\mathsf{T}} \cdot (\tilde{\mathbf{x}}[p] - \alpha[p]\mathbf{c} - \beta[p]\mathbf{s}) \tag{8}$$

$$= \sum_{p=0}^{M_1M_2-1} (\tilde{\mathbf{x}}[p] - \mathbf{H}\boldsymbol{\gamma}[p])^{\mathsf{T}} (\tilde{\mathbf{x}}[p] - \mathbf{H}\boldsymbol{\gamma}[p]) \tag{9}$$

where

$$\mathbf{c} = \left[1 \ \cos(2\pi f_0 T_s) \ \ldots \ \cos(2\pi f_0 (N-1) T_s)\right]^{\mathsf{T}}$$
$$\mathbf{s} = \left[0 \ \sin(2\pi f_0 T_s) \ \ldots \ \sin(2\pi f_0 (N-1) T_s)\right]^{\mathsf{T}}$$

are vectors of size $N \times 1$ associated with the cosine and sine components over time. In (9), the parameters $\alpha[p]$ and $\beta[p]$ and the vectors $\mathbf{c}$ and $\mathbf{s}$ are grouped by defining: $\boldsymbol{\gamma}[p] = [\alpha[p] \ \beta[p]]^{\mathsf{T}}$, with size 2×1, and the matrix $\mathbf{H} = [\mathbf{c} \ \mathbf{s}]$, with size $N \times 2$.

A formulation in terms of a simple linear model [5], [8], can be obtained using a suitable notation which groups the involved vectors and matrices as

$$\bar{\mathbf{x}} = \left[\tilde{\mathbf{x}}[0]^{\mathsf{T}} \ \tilde{\mathbf{x}}[1]^{\mathsf{T}} \ \cdots \ \tilde{\mathbf{x}}[M_1M_2-1]^{\mathsf{T}}\right]^{\mathsf{T}},$$
$$\mathbf{Z} = \begin{bmatrix} \mathbf{H} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{H} \end{bmatrix} \text{ and}$$
$$\mathbf{d} = \left[\boldsymbol{\gamma}[0]^{\mathsf{T}} \ \boldsymbol{\gamma}[1]^{\mathsf{T}} \ \ldots \ \boldsymbol{\gamma}[M_1M_2-1]^{\mathsf{T}}\right]^{\mathsf{T}}$$

where $\bar{\mathbf{x}}$, $\mathbf{Z}$ and $\mathbf{d}$ have size $N(M_1M_2) \times 1$, $N(M_1M_2) \times 2M_1M_2$ and $2M_1M_2 \times 1$, respectively. Equation in (9) can now be expressed in the form

$$J(\boldsymbol{\theta}') = (\bar{\mathbf{x}} - \mathbf{Z}\mathbf{d})^{\mathsf{T}} (\bar{\mathbf{x}} - \mathbf{Z}\mathbf{d}). \tag{10}$$

Following the classical theory of estimation for multiple parameters in linear models [5], the function (10) can be minimized over $\mathbf{d}$ for

$$\hat{\mathbf{d}} = \left(\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\right)^{-1} \mathbf{Z}^{\mathsf{T}}\bar{\mathbf{x}} \tag{11}$$

so that its minimum with respect to $\mathbf{d}$ is

$$J(f_0) = \left(\bar{\mathbf{x}} - \mathbf{Z}\hat{\mathbf{d}}\right)^{\mathsf{T}} \left(\bar{\mathbf{x}} - \mathbf{Z}\hat{\mathbf{d}}\right) = \bar{\mathbf{x}}^{\mathsf{T}} \left(\mathbf{I} - \mathbf{Z}\left(\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathsf{T}}\right)\bar{\mathbf{x}} \tag{12}$$

where $\mathbf{I}$ is the identity matrix, $\hat{\mathbf{d}}$ in (11) has been used and the dependence of $J(\cdot)$ on the remaining variable f_0 has been emphasized. This optimization is effective for $\mathbf{d}$, which includes only information relative to the amplitudes $\mathbf{a}$ and the phases $\boldsymbol{\phi}$; in order to obtain the estimation of f_0, the last equation has to be minimized over f_0 or, equivalently, maximized over the term

$$\bar{\mathbf{x}}^{\mathsf{T}}\mathbf{Z}\left(\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathsf{T}}\bar{\mathbf{x}}. \tag{13}$$

After the maximization of (13), the estimation of the parameters in $\hat{\boldsymbol{\theta}}'$ is obtained, from which the parameters $\mathbf{a}$, f_0 and $\boldsymbol{\phi}$ can be computed.

A proper approximation and simplification of the matrix $\mathbf{Z}\left(\mathbf{Z}^{\mathsf{T}}\mathbf{Z}\right)^{-1}\mathbf{Z}^{\mathsf{T}}$ can lead to an *approximate* ML estimator. Following [5], [7], [9], the ML estimator of the frequency f_0 can be obtained by maximizing the periodogram $I[k]$ over the overall p positions. More precisely, this can be obtained as

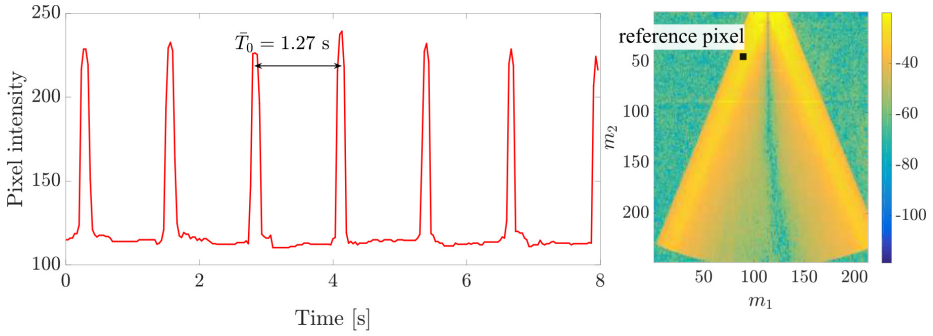
$$\hat{f}_0 = \frac{f_s}{N} \underset{k_{\min} \le k \le k_{\max}}{\arg\max} \ I[k] \tag{14}$$

where, assuming regular periodicity, the $\arg\max$ search is limited to the discretized frequencies set $[k_{\min}; k_{\max}]$, related to the real frequencies $f_{\min} = \frac{k_{\min}}{N} f_s$ and $f_{\max} = \frac{k_{\max}}{N} f_s$, and the periodogram is defined as

$$I[k] = \frac{2}{N} \sum_{p=0}^{M_1M_2-1} \left| \sum_{n=0}^{N-1} x[p,n]\, e^{-j2\pi \frac{k}{N} n} \right|^2. \tag{15}$$

This simplified estimator is approximately ML if the real frequency f_0 is not close to 0 or $f_s/2$, only.

 ISBN: 978-1-61208-548-7



(a) (b)

Figure 2. Example of periodicity analysis for the video of a pendulum. In (a) the intensity variations on the reference pixel and (b) the estimation of $\mathrm{vec}^{-1}(\mathbf{a})$ i.e., the amplitudes at the position of the various pixel.

III. Applications

In this section, the performance of the ML approach is discussed, showing its effectiveness in the extraction of periodic features. In particular, we discuss the importance of estimating the fundamental frequency f_0 in the framed image variations and the capability of localizing them inside the frame by the estimation of local amplitudes $\mathbf{a}$. The first feature is attractive, because it may be very useful in several tasks that involve the monitoring of some events or movements related to a periodic variation. The second one is equally important, since the localization of such variations may be a key feature to increase computational efficiency in some applications or video signal analysis for surveillance purposes.

In order to show the efficiency and the simplicity of the ML approach, three examples in different scenarios are reported, describing the capabilities of the approach and focusing on its properties in each application. Specifically, the examined scenarios are related to:

1) analysis of physical oscillations
2) analysis on movements of athletes and people doing gymnastic activity
3) monitoring of vital signs in animals and humans.

A. Physical Oscillations

As a first application example, we analyze the periodicity of the oscillations of a physical pendulum captured by a still camera positioned in front of the pivot. This example demonstrates the effectiveness of ML estimation on multidimensional signals.

The recorded video sample, where few frames were preliminarily depicted in Figure 1, shows an oscillating plank, with the pivot hooked on a border of a desk. Selecting a proper reference pixel, it is possible to show the intensity variations over time connected with the periodic passage of the pendulum on the involved pixel.

In Figure 2(a), these variations over the time dimension of the multidimensional signal are displayed: the peaks inside the signal correspond to the passage of the white pendulum on the reference position, which has higher intensity values than the dark background. By measuring the distance between the peaks, it is possible to estimate the average rate of the oscillation: by inspection of the signal in Figure 2(a), an average oscillation time of $\bar{T}_0 = 1.27$ Hz can be obtained, corresponding to a fundamental frequency $f_0 = 0.787$ Hz: this value is used as reference and can be compared with the estimate extracted by the video estimation system. Applying the approximate ML approach (14) on the considered sample video, a frequency $\hat{f}_0 = 0.77$ Hz is estimated, with a relative estimation error equal to 2.16%.

The influence of the periodicity on every pixel is computed by the estimation of $\hat{\mathbf{a}}$. Using (11), the amplitudes are obtained and shown in Figure 2(b), where the results are shown as an image with size equal to that of original video frames. It can be noticed that in the area directly below the pivot the estimated amplitudes have lower values: this effect is due to the fact that in this area the pixel intensities are stressed by variations with a rate doubled with respect to the fundamental one. Differently, the areas on the left and right of the axes of the pivot are mainly affected by the fundamental periodicity: therefore, the intensity of the estimated amplitude is higher. It is remarked that the estimated amplitudes are reported in a logarithmic scale, with the purpose to enhance and make more visible the difference between the various areas.

B. Athlete Monitoring

As a first realistic application example, the scenario of monitoring of physical activity made by people or athletes is presented. In fact, many physical exercises involve periodic movements or repetition of a single gesture: examples of these movements are given by weight-lifting, sit-ups and stretches. These repetitive movements are expected to involve specific body parts without a global motion of the gymnast, as they are performed on a fixed position.

To show the effectiveness of the ML approach in this environment, we consider a video sample of a man doing a series of push-ups. The duration of the video sample is about 26 s, it is recorded with a frame rate $f_s = 30$ Hz and has a frame size of 516×216 pixels. By visual inspection, the man was able to do about 19 push-ups during the whole video, with an approximate average frequency of $\bar{f}_0 = 0.73$ Hz. In Figure 3(a), a sample frame of the considered video is shown.

Initially, the analysis of the local amplitudes is clarified: after an estimate by video processing of the fundamental frequency with an average push-up rate of $\hat{f}_0 = 0.725$ Hz, the parameter $\mathrm{vec}^{-1}(\hat{\mathbf{a}})$ is computed and shown in Figure 3(b). As in the example of the pendulum described in Subsection III-A, pixels with higher value are those mainly affected by the periodic motion of the push-ups. On the other side, the background has lower intensity value, since pixel variations are modified only by random motion on the scene or noise. In Figure 3(b), the estimated amplitudes are reported in logarithmic scale. It

 ISBN: 978-1-61208-548-7

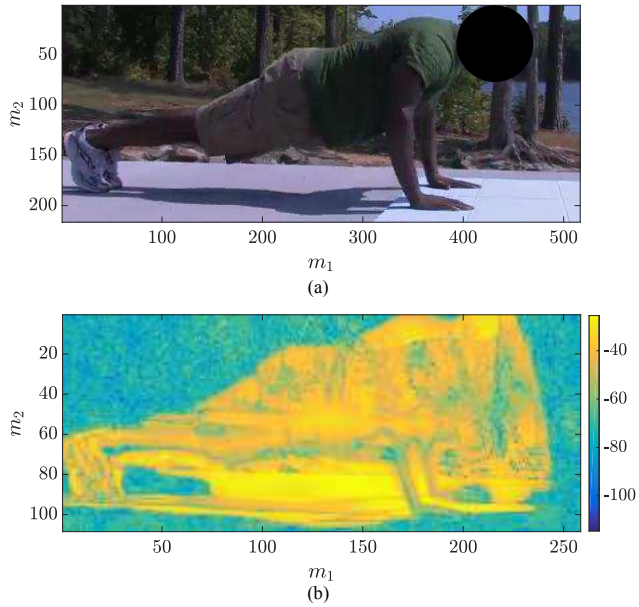


Figure 3. Example of athlete monitoring: (a) sample frame of a man doing push-ups and (b) amplitude estimation for $\hat{f}_0 = 0.725$ Hz.

is clear that this analysis can be used to localize the periodic motion and, as an example, create a mask for further video processing algorithms.

Afterwards, a deeper analysis of the fundamental frequency estimation is performed. The repetition times of the athlete doing push-ups were measured by the use of a stopwatch and computing a curve fitting of the evolution of the push-up rates over time. The video was analyzed with temporal windows of $NT_s = 10$ s and an overlapping parameter of 90%, obtaining an estimation of the fundamental frequency for every second. In Figure 4, the stopwatch reference compared with the rates estimated by applying the ML approach proposed in Section II is shown; the quality of the estimation is clear, exhibiting also a pattern similar to that of the original rates.

As further evidence of the effectiveness of the proposed approach in the estimation of the periodic features of the video signal, the error on the estimation of f_0 is also reported. Considering the results shown in Figure 4, a Root Mean Squared Error (RMSE) of 0.0128 Hz is obtained, which, normalized to the average value of the reference, gives an average relative error of 1.8%.

C. Monitoring of Vital Signs

As last example of the reliability of the ML approach for periodic feature extraction, an application in the biomedical scenario is proposed. In particular, monitoring of vital signs is a key tool to asses the health condition of a patient. Recent studies [9]–[13] report that some of the vital signs, such as hearth and respiratory rates, can be evaluated by contactless systems employing video cameras and multidimensional signal processing. Among vitals signs, the Respiratory Rate (RR) plays a very important role as indicator of the health of a patient. It is now demonstrated that the proposed ML approach can be used for both tasks of estimating the RR of a framed patient and localizing the areas mainly affected by respiratory movements. This last feature may be very useful in order to reduce computational complexity of video processing-based

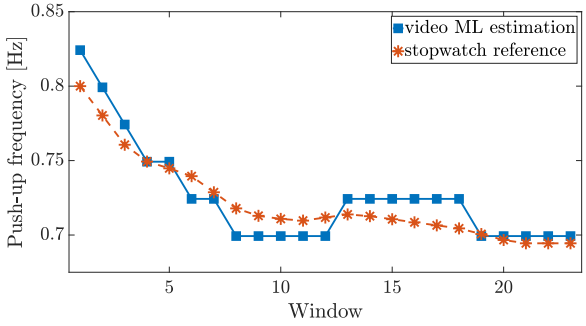


Figure 4. Performance evaluation in the estimation of the fundamental frequency f_0, related to the push-up rate of an athlete.

algorithms by localizing Regions of Interests (ROI), as shown in [9].

A first test is performed on monitoring the respiration of a sleeping cat. The animal was completely still and breathing with a constant RR $f_0 \in [0.28, 0.35]$ Hz, obtained by a chronograph. The RR reference measurements were obtained by live inspection during video recording by careful observation of the animal. These measurements can be easily converted to breaths per minute (bpm) if desired. In Figure 5(a), a sample frame of the video sequence is shown: the video signal has a total duration of 1 min and 13 s with a sampling rate of 15 Hz and a camera resolution of 320×240 pixels.

In Figure 5(b), the likelihood function $J(f_0)$ used for the estimation of the fundamental frequency is shown. The periodicity related to breathing movements obtained by processing the variation of pixel intensity is clear. Taking the $\arg\max$ of the likelihood function, the frequency $\hat{f}_0 = 0.3$ Hz can be estimated, according to the frequency range used as reference. As discussed in Section II, after the estimation of the fundamental frequency, the parameter $\mathrm{vec}^{-1}(\hat{\mathbf{a}})$ can be computed. In Figure 6, the estimated pixel-wise amplitudes are shown. Higher values are obtained in the pixel positions mainly involved in breathing movements that are near the chest and the abdomen of the cat. By selecting this area as a possible ROI, it is feasible to develop algorithms that are robust against possible large random movements [9], excluding other areas that are involved in useless movements or random noise.

As a last test, relying on the work presented in [9], the estimation of the RR on a real newborn patient and the localization of breathing areas are performed. The video was recorded in the University Hospital of Parma, by video cameras with resolution of 720×576 and sampling rate $f_s = 25$ Hz, with an overall duration of 3 min and 3 s. As in the push-up example, the RR is here estimated over time and compared with the pneumographic reference, the gold-standard system for monitoring of respiration mainly used by clinicians. Using windows of analysis of length $NT_s = 20$ s and an overlapping parameter of 90% (i.e., with RR estimation obtained every 2 s), the comparison between the reference and the estimation by video processing is depicted in Figure 7. Excluding the first five windows, where the algorithm has startup issues, the correspondence of the estimated RRs with the pneumographic ones is very good.

As described in Subsection III-B, also for this example a thorough analysis is performed, reporting the RMSE and the average relative estimation error on the RR. A RMSE equal to

 ISBN: 978-1-61208-548-7

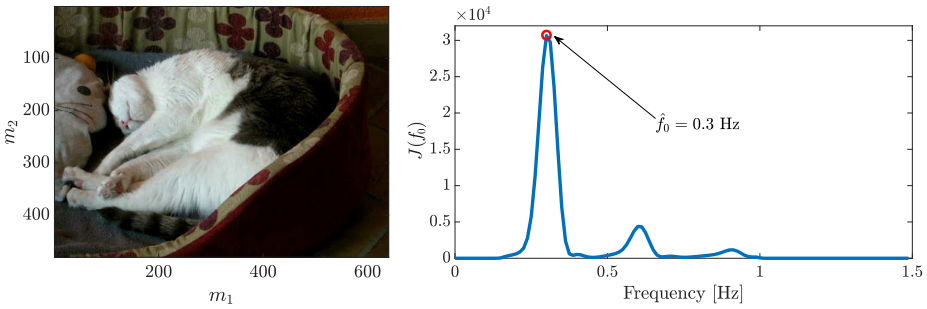


Figure 5. Monitoring of a sleeping cat: (a) frame sample from the video recording and (b) the likelihood function $J(f_0)$ for the estimation of the RR.

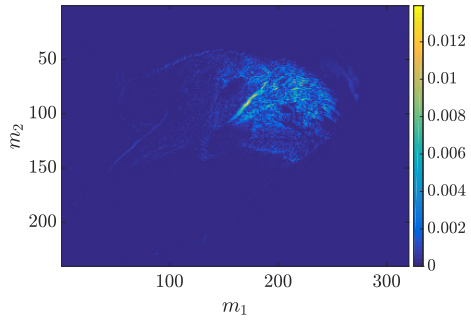


Figure 6. Estimated amplitudes for the video example of a sleeping cat: maximum values of $\text{vec}^{-1}(\mathbf{a})$ can be noticed near the chest of the animal.

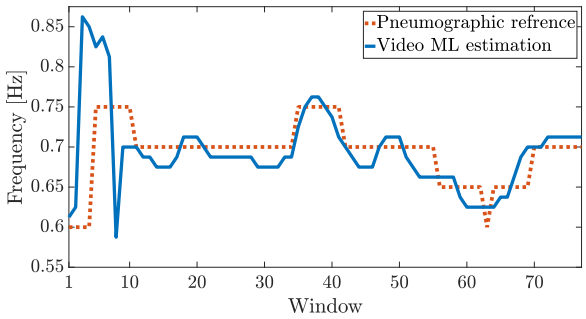


Figure 7. Estimation of the RR in monitoring of a newborn: comparison of the estimation from video signals and the pneumographic device.

0.051 Hz is obtained with an average relative error of 7.5% An in-depth analysis on the performance for RR estimation of the technique introduced in Section II is beyond the scope of this paper. The interested reader is referred to [9]. Nonetheless, the presented results highlight the usefulness of the ML approach applied to multidimensional video signals for the extraction of periodic features.

IV. Conclusion

In this paper, we proposed a method for the extraction of periodic features in video signals. Under the assumptions of still camera and that the framed subject is not affected by translation or superposed motion, we introduced a model of periodicity in multidimensional signals; then, we applied the ML criterion for the estimation of the periodic features of interest. Finally, we demonstrated the effectiveness of this approach, showing three different application examples: monitoring of physical oscillations, athlete movements and vital signs. The advantage in the localization of periodic variations and estimation of the fundamental frequency has been demonstrated by comparing the obtained results with suitable reference values.

References

[1] A. Briassouli and N. Ahuja, "Extraction and analysis of multiple periodic motions in video sequences," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 7, July 2007, pp. 1277–1261.

[2] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis, and applications," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, Aug. 2000, pp. 781–796.

[3] G. M. Kouamou Ntonfo, G. Ferrari, R. Raheli, and F. Pisani, "Low-complexity image processing for real-time detection of neonatal clonic seizures," IEEE Trans. Inf. Technol. Biomed., vol. 16, no. 3, May 2012, pp. 375–382.

[4] H. L. Van Trees, Detection, Estimation, and Modulation Theory (Part I), 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 2001.

[5] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory. Upper Saddle River, NJ, USA: Prentice Hall, 1993, vol. 1.

[6] L. Cattani et al., "Monitoring infants by automatic video processing: A unified approach to motion analysis," Comput. Biol. Med. (Elsevier), vol. 80, Jan. 2017, pp. 158–165.

[7] N. Patwari, J. Wilson, S. Ananthanarayanan, S. Kasera, and D. Westenskow, "Monitoring breathing via signal strength in wireless networks," IEEE Trans. Mobile Comput., vol. 13, no. 8, Aug. 2014, pp. 1774–1786.

[8] C. M. Bishop, Pattern Recognition and Machine Learning, 1st ed. New York, NY, USA: Springer-Verlag, 2006.

[9] D. Alinovi, G. Ferrari, F. Pisani, and R. Raheli, "Respiratory rate monitoring by maximum likelihood video processing," in Proc. IEEE Int. Symp. Signal Process. and Inf. Technol. (ISSPIT), Limassol, Cyprus, Dec. 2016, pp. 172–177.

[10] D. Alinovi, L. Cattani, G. Ferrari, F. Pisani, and R. Raheli, "Spatio-temporal video processing for respiratory rate estimation," in Proc. IEEE Int. Symp. Med. Meas. and Applicat. (MeMeA), Turin, Italy, June 2015, pp. 12–17.

[11] L. Tarassenko et al., "Non-contact video-based vital sign monitoring using ambient light and auto-regressive models," IOP Physiol. Meas., vol. 35, no. 5, May 2014, pp. 807–831.

[12] R. Janssen, W. Wang, A. Moço, and G. de Haan, "Video-based respiration monitoring with automatic region of interest detection," IOP Physiol. Meas., vol. 37, no. 1, Jan. 2016, pp. 100–114.

[13] C.-W. Wang, A. Hunter, N. Gravill, and S. Matusiewicz, "Unconstrained video monitoring of breathing behavior and application to diagnosis of sleep apnea," IEEE Trans. Biomed. Eng., vol. 61, no. 2, Feb. 2014, pp. 396–404.

 ISBN: 978-1-61208-548-7