

Real-Time Temporal Superpixels for Unsupervised Remote Photoplethysmography

Serge Bobbia*, Duncan Luguern*, Yannick Benezeth*
Keisuke Nakamura**, Randy Gomez** and Julien Dubois*

*LE2I EA7508, Univ. Bourgogne Franche-Comté (UBFC), Dijon, France

**Honda Research Institute Japan Co., Ltd., 8-1 Honcho, Wako-shi, Saitama, Japan

Abstract

Segmentation is a critical step for many computer vision applications. Among them, the remote photoplethysmography technique is significantly impacted by the quality of region of interest segmentation. With the heart-rate estimation accuracy, the processing time is obviously a key issue for real-time monitoring. Recent face detection algorithms can perform real-time processing, however for unsupervised algorithms, i.e. without any subject detection based on supervised learning, existing methods are not able to achieve real-time on regular platform. In this paper, we propose a new method to perform real-time unsupervised remote photoplethysmography based on efficient temporally propagated superpixels segmentation. The proposed method performs the segmentation step by implicitly identifying the superpixel boundaries. Hence, only a fraction of the image is used to perform the segmentation which reduces greatly the computational burden of the process. The segmentation quality remains comparable to state of the art methods while computational time is divided by a factor up to 8 times. The efficiency of the superpixel segmentation allow us to propose a real-time unsupervised rPPG algorithm considering frames of 640x480, RGB, at 25 frames per second on a single core platform. We obtained real-time processing for 93% of precision at 2.5 beat per minute using our inhouse video database.

1. Introduction

The photoplethysmography method (PPG) is used in many devices as pulse oxymeter or fitness smart watch. The blood volume variation measured this way is primarily used to determine the heart rate. Moreover, oxygen saturation or breath rate can also be determined with this measure. The PPG measurement is usually performed with contact to the skin. A light source illuminates the tissue and the small variations of the reflected or transmitted lights are measured

with a photodetector [27]. However, this method requires a stable contact between the sensors and the skin which tends to make this method not suitable for damaged skin or if the contact cannot be maintained steadily. The motion and the pressure applied on the sensor can also affect the signal waveform [29].

Based on the computer vision advances, recent works demonstrate that cardiovascular activities can be monitored using back-scattered ambient light [27]. It requires no expensive or specific hardware as a standard camera sensor (smart-phone or web-cam) can be used to perform the measure. Most methods share a common pipeline-based framework. First, a region of interest (ROI) is selected and tracked over frames. Second, the RGB informations are averaged over the entire ROI and combined to constitute the temporal traces which is then filtered. Finally, the heart-rate can be estimated, based on the frequency decomposition of the remote photoplethysmographic (rPPG) signal.

The pulsatile information is mixed with the reflected light and shadow casting over the skin due to movement. This mixed signal is then captured by the camera. It suggests that blind source separation methods can be used to extract rPPG information from the mixed signal. The RGB mixture can then be linearly combined in order to maximize independence. Hence, Independent Component Analysis (ICA) has been used in several works [16] or [24] along with Principal Component Analysis (PCA) in order to extract the signal which maximize the variance [18]. These methods require no knowledge of the plethysmographic signal. Otherwise, prior information of the signal has lead to the development of several methods. In [21] and [22], the authors proposed to incorporate a priori information about the periodicity of the cardiac signal into a constraint ICA algorithm or a multiobjective optimization framework to aid in the extraction of the most prominent rPPG signal. On the other hand CHROM [9], PBV [12] or POS [15] use fixed coefficients between the RGB channels to maximize

the rPPG informations while minimizing the perturbation induced by motion. Some methods have even used data-driven approach to adapt to the scene [33].

The ROI selection step is crucial to obtain a reliable signal. From earlier studies to recent works, ROI selection has evolved through several approaches. Initially manually selected [30, 26], most of the methods are using face detection [24] and tracking algorithms [28] and possibly refined with a skin pixel classification [31] to select ROI. Lately, refined ROI selection has been proposed for maximizing the information quality [35]. It leads to an higher Signal Noise Ratio (SNR) and so a more robust heart rate estimation based on prior information as it has been shown that face does not exhibit rPPG information homogeneously [17]. Indeed, cheek and forehead contribute significantly more than other areas.

For unsupervised segmentation, *i.e.* without any subject detection based on supervised learning, methods use pixel clustering to create multiple ROI which are evaluated independently to determine their contribution to the rPPG information. Wang et al. [32] use voxels to construct spatio-temporal clusters and sparse PCA method to determine the contributive voxels. This method has then been extended to segment living skin based on rPPG information recognition [34]. In [3], temporally propagated superpixels are used as the segmentation step and later merged according to their SNR. It leads to a maximization of the rPPG information. All of these methods suffer from a high computational burden. Indeed, the temporal superpixels are made using the SLIC method [2] for superpixels construction and optical flow to determine the temporal propagation which is computationally extremely expensive. In the case of voxels, spatio-temporal clustering requires too much costly operations to be performed in real-time on common PC platforms. It has been shown that the number of pixels in a ROI has a direct impact on the signal quality [4]. It means that the segmentation level of the frame has a direct impact on the signal quality [3]. As a higher superpixel number means less pixels per cluster, the RGB quantization error increases accordingly. This may be understood as the reduction of the sensor noise amplitude by a factor equal to the square root of the number of pixels used in the averaging process of the ROI [11]. However, a small number of superpixels leads to a less accurate segmentation which also affects the signal quality because of non-skin pixels [33].

This paper introduces a new real-time temporal superpixels segmentation method for remote photoplethysmography. As most of the computational burden of the previously introduced methods are linked with the segmentation task, we propose in this paper a new way of performing superpixels segmentation which requires notably less computation. Our method, called Iterative Boundaries implicit Identification for superpixel Segmentation (IBIS), significantly limits the computation burden by implicitly identifying the super-

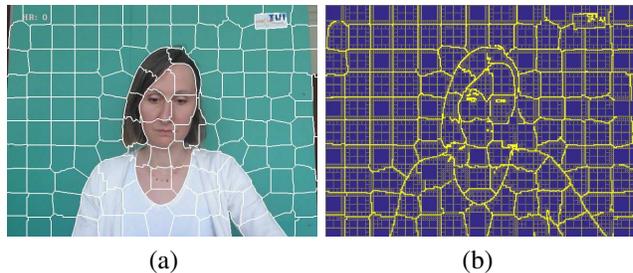


Figure 1. Example of segmentation with IBIS: (a) Segmented image, (b) Pixels of the image that are actually processed.

pixels boundaries without processing the entire image. The process can be performed up to 8 times quicker than current state of the art methods. Figure 1 shows an example of IBIS segmentation with the pixels of the image that are actually processed in order to retrieve superpixels boundaries. Moreover as the number of superpixels increases, the number of iterations decreases to guarantee a low computation burden even for a large number of superpixels. The temporal propagation is performed relatively to the modification of the consecutive images. Thus, the creation and deletion of superpixels are performed such that the number of clusters remains the same. To overcome the tedious selection of the number of superpixels, we propose an automatic segmentation level selection in order to maximize the quality of the rPPG information while guarantying a good segmentation quality.

The method is detailed in section 2. Then the experiments are presented in section 3 which introduces the dataset and the metrics used to evaluate both the superpixels method IBIS and the temporal propagation with regards to the quality of the rPPG measurement. The conclusion and discussion are proposed in section 4.

2. Method

First the proposed superpixel segmentation method is presented in subsection 2.1. Then, the temporal propagation principles are presented in subsection 2.2 and finally the rPPG heart-rate estimation is presented in subsection 2.3.

2.1. Efficient superpixels segmentation

Superpixels algorithms perform segmentation in order to group pixels in a coherent way with respect to the edges. They are used for a wide range of applications such as object tracking [14], image segmentation [19], video representation [5] or for biomedical applications [3]. The clusters are established based on various scenarios.

Graph-based methods model the relation between neighboring pixels with graph. As an example, the Normalized cuts method [25], represents pixels as nodes and link to pixel neighbors as edges. Segmentation is done by recursively minimizing a cost function on the graph. Guid-

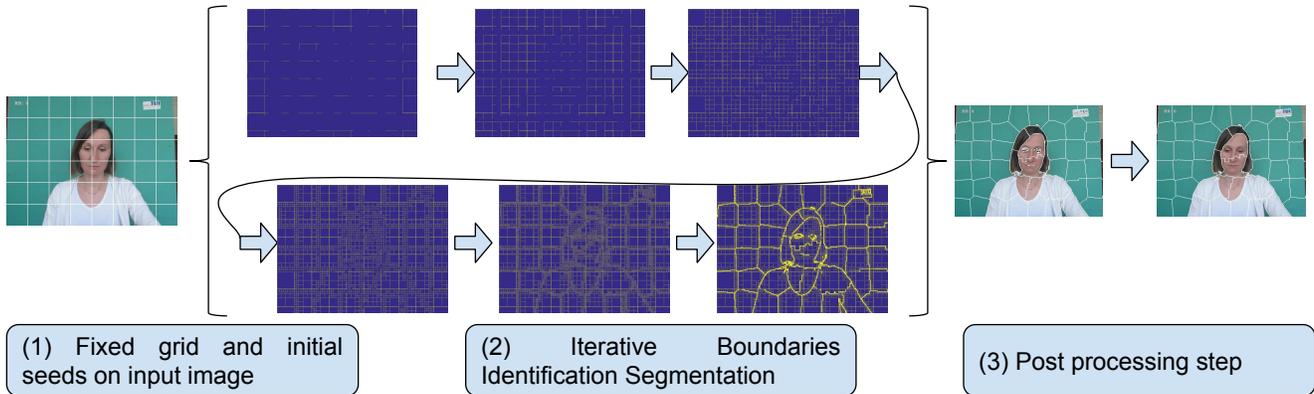


Figure 2. Overview of the method: (1) Initial seeds definition with fix grid, (2) IBIS segmentation and (3) Post processing step.

ing model [23] reduces the complexity of the segmentation based on the same approach.

Other methods based on gradient-ascent use initial regions as references and refine boundaries location to generate compact superpixels. Mean shift [6], SLIC [2] or USEQ [13] segment the image in spatially regular clusters based on the euclidean distance evaluated between superpixels and all the pixels in the image.

All the existing methods focus on increasing segmentation quality or decreasing the computational cost for the process. Our method focuses on the complexity by proposing a simple iterative way to perform the segmentation by processing a fraction of the image.

Figure 2 shows an overview of the proposed method. (1) Seeds are initialized based on a fixed grid. (2) Segmentation is performed by iterative and implicit boundaries identification on blocks of pixels decreasing recursively the size of the block. (3) A post-processing step is added to ensure superpixel contiguity and a minimal size.

2.1.1 Superpixels generation

Introduced by Archanta et al. [2], the SLIC method is one of the main contribution in the field of superpixels. It is still one of the most reliable superpixels segmentation method [1] and is still widely used and cited (e.g. [7], [8]). The clustering is done by agglomerating pixels to initially define superpixels centers, later called seeds, with respect to the spatial and chromatic euclidean distances. The chromatic space is the CIE $l^*a^*b^*$ because this color space is widely considered as perceptually uniform for small color distances [2]. The process is repeated 10 times to ensure a high segmentation quality to be reached as the seeds are updated between each iteration as 5-dimensional vectors composed of the averaged pixel coordinates x and y along with the averaged $l^*a^*b^*$ color values. Due to the simplicity of the distance computation and the quality of the resulting segmentation, we have based our procedure to assign pix-

els to the superpixels on SLIC.

As in SLIC, pixels are agglomerated following their color similarity and spatial proximity in the image plane. Pixels in the image are associated with the nearest seed according to the sum of the color distance D_{lab} and the spatial distance D_{xy} . The total distance D between the i^{th} pixel and the k^{th} seed (with $k = [0, K]$ and K the number of superpixels) is defined by $D = D_{lab} + \theta * D_{xy}$ with

$$D_{lab} = \|(l, a, b)_i - (l, a, b)_k\|, \quad (1)$$

$$D_{xy} = \|(x, y)_i - (x, y)_k\|. \quad (2)$$

$\|\cdot\|$ is the Euclidian norm and θ is a compacity factor defined by $\theta = 1/c^2$ with c a parameter fixed by the user. This operation is the most computationally expensive. Indeed considering each iteration, every pixel has to be evaluated for all possible seeds. The cost of conventional approaches cannot be reduced and varies according to the number of pixels in the image and according to the number of superpixels specified by the user. Obviously, the search area should be limited in order to reduce the computational burden and as in SLIC, pixels are compared to the adjacent seeds within a range defined by the user. This way pixels are evaluated to a subset of seeds value only considering their spatial repartition.

The IBIS method is similar to the SLIC procedure presented above. However, in addition to limiting the computation to a locality considering a subset of seeds for evaluation, we introduce a new approach to limit significantly the quantity of pixels evaluated in the image. This procedure is presented in the next subsection.

2.1.2 Boundaries identification

Based on results generated by state of the art methods as SLIC [2], SEEDS [10] or USEQ [13], superpixels can be defined by common properties. Superpixels should be regular, contiguous and have a minimal size. The output of the

segmentation is an assignation of every pixels to a super-pixel. We propose to use a set of square block to iteratively identify the boundaries between superpixels.

The image is first divided into blocks using a regular grid. A block is said to be *coherent* if a subset of pixels within that block, called the *processed pixels*, are assigned to the same superpixel. In that case, all pixels of this block are assigned to the corresponding superpixel. On the other hand, as soon as the *processed pixels* of a block are assigned to different superpixels, the process is stopped then the block is divided into four smaller blocks and the process is repeated iteratively. With this strategy, it is possible to drastically limit the number of pixels that are evaluated. Actually, only the pixels located on the border of a block are evaluated to determine whether a block is *coherent* or not. By doing this, we determine iteratively blocks that contain the edges of the image and boundaries of superpixels. The procedure is further optimized by evaluating only one pixel over two that is on the block boundary. These pixels are called the *processed pixels* and represent only a fraction of the image (cf. Figure 1).

The block size τ_i for the i^{th} iteration is defined by:

$$\tau_i = 2^{\alpha-i+1} + 1 \quad (3)$$

with

$$\alpha = \lfloor \log_2(\sqrt{N/K}) - 1 \rfloor \quad (4)$$

$i \in [1, \alpha]$, $\sqrt{N/K}$ the average size of the superpixels and N the number of pixels in the image.

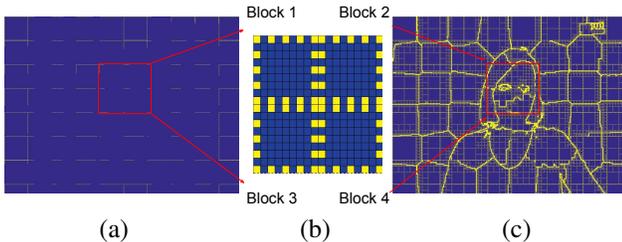


Figure 3. Superpixels boundaries are identified by evaluating one pixel over two on the block boundary. (a) Pixels of the top-level blocks that are actually processed. (b) Top-level blocks are independent and (c) Iterative Boundaries Identification.

Top-level blocks (at first iteration) are completely independent and can therefore be evaluated in parallel, which allows for massive parallelization of the segmentation process. For this purpose, it is necessary that the blocks are formed so that the boundaries do not overlap. The boundaries of two top-level neighboring blocks are simply contiguous as shown in Figure 3.

For the last iteration α (corresponding to a block size of 3, i.e. $\tau_\alpha = 3$), if the *processed pixels* are not assigned to the same superpixel, i.e. the block is not *coherent*, all the pixels within that block are individually evaluated so that the superpixels adhere precisely to the image edges.

2.1.3 Seeds update

In the initial SLIC method, the seeds update allows for an increase of the quality of the *Boundary Recall* as the seeds agglomerate more homogeneous pixels based on their color similarity and proximity in the image plane. In our approach, seeds are also updated after each iteration, as an average of all the pixels assigned to the corresponding superpixel. If no pixel is assigned to this superpixel, its seed remains unchanged.

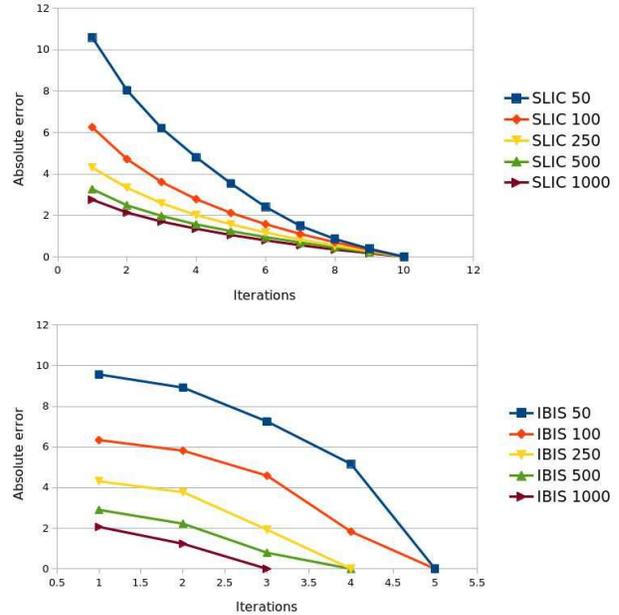


Figure 4. Evolution of the absolute difference between the seed value at a current iteration and the seed value after convergence with respect to the number of iterations for SLIC (top) and IBIS (bottom). The absolute error is the averaged error on the 5-dimensional seeds values for all superpixels.

Figure 4 shows the evolution of the absolute error, between the seed value at a current iteration and the seed value after convergence for both SLIC and IBIS methods. For compactness, 5D seed vector values are averaged.

As the number of superpixels increases, the quantity of pixels per superpixel decreases and consequently the variations of the seeds values will be smaller. For SLIC, the convergence is reached after 10 iterations regardless of the number of superpixels as it is fixed in the algorithm. For the IBIS method, the number of iterations is dependent of the number of block sizes which is determined by the number of superpixels (cf. Eq. 3). The shape difference of the curves is due to the way seeds are updated in both methods. For SLIC, all pixels in the image are assigned to the nearest superpixel. On the contrary, for IBIS, through the iterations and the block-size decreasing accordingly, a larger number of blocks will be *coherent* as the probability of hitting a boundary will decrease with the masks size getting smaller.

The post-processing step is the same as the SLIC method

where a mask is applied along all the image to ensure the contiguity and a minimal size for all superpixels.

2.2. Temporal propagation

In order to perform the rPPG measurement as proposed in [3], we have to consider the temporal propagation of the seeds value. Hence, each temporal superpixels provides an instantaneous averaged chromatic value as long as the best possible inheritance. Two superpixels are said to be adjacent if they share the same boundary. Parent superpixels for the frame $t - 1$ are defined by the minimal euclidean distance for a seed value at time t with the seed and adjacent seeds at time $t - 1$. As pixels can only be exchanged between adjacent superpixels, this method ensures to construct the most coherent possible RGB traces.

The temporal propagation of the superpixels is done with respect to the pixels that have been modified between two consecutive frames. Only pixels that could possibly have been exchanged with adjacent superpixels are processed. Such pixels are considered if the average absolute chromatic difference between spatially identical pixels is superior to an empirically selected threshold. It allows for an additional computational cost reduction as long as a high stability of the background superpixels in the scene.

The post processing step ensures the superpixel contiguity as long as deleting the superpixels that are too small, which contains a number of pixels inferior to four times the average size of the image superpixels, *i.e.* $(\sqrt{N/K})/4$. The number of clusters should remain stable in the scene and their size kept under control over the frames. Hence, the k largest superpixels are divided in two superpixels, with k the number of clusters deleted in the previous step. Splitting superpixels generates two new seeds only separated by their spatial coordinates for one pixel in both dimension. Once the seed value are updated, only pixels contained in deleted superpixels and the split ones are processed again to update the segmentation.

2.3. Heart-rate estimation

As in [3], the input video frames are decomposed into several temporal superpixels from which the pulse signals are extracted. CHROM method is used for each temporal superpixel and a pulsatility measure is then used to merge the pulse traces and estimate the photoplethysmogram signal. For each temporal superpixel, the RGB traces are band-pass filtered from 40 to 240 bpm and RGB channels are linearly combined according to the CHROM [9] method. It consists on projecting the RGB traces onto two orthogonal vectors defined as :

$$\begin{aligned} X_i(t) &= 3y_i^R(t) - 2y_i^G(t), \\ Y_i(t) &= 1.5y_i^R(t) + y_i^G(t) - 1.5y_i^B(t). \end{aligned} \quad (5)$$

The pulse signal S_i , for the i^{th} temporal superpixel, is calculate with $S_i(t) = X_i(t) - \alpha Y_i(t)$ where $\alpha_i = \sigma(X_i)/\sigma(Y_i)$. The rPPG signals are fused based on their quality, estimated with the SNR value as:

$$SNR_i = 10 \log_{10} \frac{\int_{f_1}^{f_2} h_{signal}^i(f) |\mathcal{F}\{S_i(t)\}|^2 df}{\int_{f_1}^{f_2} h_{noise}^i(f) |\mathcal{F}\{S_i(t)\}|^2 df} \quad (6)$$

where $\mathcal{F}\{S_i(t)\}$ is the Fourier transform of the rPPG signal of the i^{th} temporal superpixel. The double-step function h for the two first harmonic defined as:

$$\begin{aligned} h_{signal}^i(f) &= [\delta(f - f_0^i) + \delta(f - 2f_0^i)] * \prod(\pm f_r) \\ h_{noise}^i(f) &= 1 - h_{signal}^i(f) \end{aligned} \quad (7)$$

with δ the Dirac delta function, f_0^i the fundamental frequency, convoluted with the *rect* function, noted as \prod of half-width f_r . SNR_i will be high for skin superpixels and low for background ones.

The heart-rate is estimated as the fundamental frequency of the rPPG signal $S(t)$ which is obtained by a weighted average of all the pulse signals $S_i(t)$, *i.e.* $S(t) = \sum_{i \in K} S_i(t) w_i$ with w_i :

$$w_i = \frac{10^{SNR_i}}{\sum_{i \in K} 10^{SNR_i}} \quad (8)$$

The weights are normalized and in order to favor pulsatile signals. They are defined with 10 time the $\log^{-1}(x)$ function (*i.e.* 10^x).

In addition, in order to take into account the need to have superpixels with well-defined boundaries and to have superpixels large enough to reduce quantization noise, the adjacent superpixels with similar color (evaluated with the euclidian chromatic distance) are also considered when the pixels are spatially averaged. By doing this, the size of the superpixels can be small enough to maintain the quality of the superpixel boundaries, and the spatial average considers a large number of pixels to reduce the impact of the quantization noise. The number of superpixels is defined by $\nu * \beta$ with β the average number of adjacent superpixels (typically $\beta = 6$) and ν the optimal superpixel resolution.

3. Experiments

First the superpixel evaluation is done in 3.1 we have compared the quality of the IBIS segmentation with state-of-the-art methods. Then the rPPG accuracy is evaluated in subsection 3.2. Performance tests are run with codes provided by the authors and are performed on the same platform, which runs an Intel i7 4790@3.6 GHz. First we describe the datasets used to evaluate the proposed method.

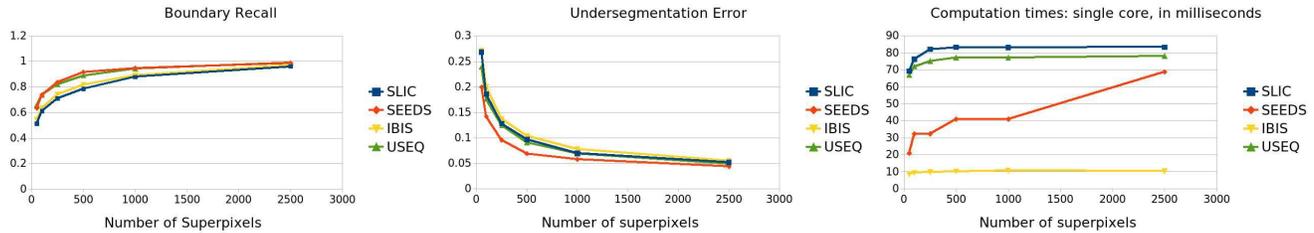


Figure 5. Comparisons of superpixels method IBIS with state of the art methods SLIC, SEEDS and USEQ for boundary recall, undersegmentation error and computation time metrics.

Then we present the evaluation metrics and finally we discuss on the results. Each metric result is the average over the entire dataset in both experiments.

3.1. Superpixels segmentation evaluation

Superpixels evaluation is done with comparison to three state-of-the-art methods SLIC [2], SEEDS [10], and USEQ [13]. Each implementation was provided by the author on their website. The IBIS and temporal IBIS codes are available on the project web-pages^{1, 2}.

3.1.1 Datasets and Metrics

For the superpixels quality evaluation, the Berkeley segmentation benchmark BSDS500 dataset was selected [20]. It contains 500 manually labeled segmented images. This dataset is one of the most commonly used for superpixels evaluation.

The following metrics are applied for all the methods and used for comparison:

- **Boundary Recall (BR)**, represents the correctness of adhering the true boundaries of objects.
- **Under segmentation Error (UE)**, measures the superpixel overlapping with multiple objects by the percentage of pixels that leak from the ground truth boundaries.
- **Computation time**, was measured with the help of the `std::chrono::duration` from the C++ standard library. The measured times are the segmentation processing times plus the post processing step. The initial tasks of defining the seeds and the initial grid are not taken into account as multiple images with the same resolution can be processed with the same parameters. For fair comparisons, times are given for forced single core execution as SLIC and SEEDS methods does not gain in performance running on multi-core platform with the implementations given by the authors.

Computation times are only indicative of the relative complexity. Indeed, as the complexity of the four methods are all on $O(N)$, times are given relatively of the level of optimizations of the integration and are relative to the hardware platform. Nevertheless, the observed performance gap

between the methods, highlights the relative complexity difference.

3.1.2 Results

In Figure 5, the segmentation quality of three state-of-the-art methods and IBIS are compared.

The SEEDS method has higher BR and a lower UE which make it the best segmenting in our test. Nevertheless, our method is consistently as good as the SLIC method, even slightly better considering the BR. Which was expected as the superpixel generation procedure of SLIC and IBIS is similar.

Considering that only a fraction of an image is considered for the segmentation, the IBIS method performs remarkably well. On the performance side, up to 8 times faster than the original SLIC method for similar quantitative results, our method is much faster than the others. Despite the fact that its run-time depends on the number of superpixels, the SEEDS method performs well once again. As expected, for the SLIC, USEQ and IBIS method, the computation time considering the number of superpixels increase much slower due to the fact that the search space for the superpixels assignment is limited to a predefined range.

3.2. Heart rate estimation evaluation

The proposed method has been compared to the implementation done in [3] that proposed similar video decomposition into several temporal superpixels from which the pulse signals are extracted. However, in [3] the temporal segmentation is based on TSP [5] which is based on SLIC and is significantly slower than IBIS. The heart rate is estimated on a 20 seconds length window. The superpixels resolution is set to 150 for non averaged seeds value and 900 for seeds values averaged over the adjacent superpixels.

3.2.1 Datasets and Metrics

The rPPG quality measurement is evaluated on the UBFC-RPPG dataset presented in [3] and publicly available, composed of 43 videos, where each video is synchronized with a pulse oximeter finger clip sensor (Contec Medical CMS50E) for the ground truth. Each video is about 2 minutes long and recorded with a low cost webcam (Logitech

¹<https://github.com/xapha/IBIS>

²https://github.com/xapha/IBIS_Temporal

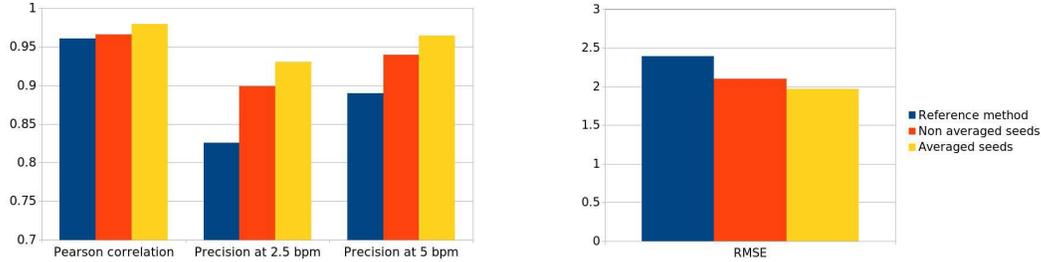


Figure 6. Comparison of the heart rate estimation metrics obtained with the reference method [3] and the proposed method with and without considering adjacent superpixels during the ROI pixel averaging.

C920 HD pro) at 30 frames per second with a resolution of 640×480 in uncompressed 8-bits RGB format.

The following metrics are applied for all the methods and used for comparison:

- **Pearson correlation factor** r is the correlation between heart rate estimated from the rPPG signal and the heart rate estimated from the PPG reference signal.
- **Root mean square error (RMSE)** is the quadratic error calculated between the measured value and the ground truth.
- **Precision at 2.5 and 5 bpm.** This metric represents the percentage of estimations where the absolute error is under a threshold (2.5 or 5 bpm)

The estimations at 2.5 and 5 bpm are expressed within the range $[0; 1]$ as well as the Pearson correlation factor metric. The *RMSE* is unit-less as it is the quadratic error measurement.

3.2.2 Results

Figure 6 presents the comparison of the heart rate estimation metrics obtained with the reference method [3] and the proposed method with and without considering adjacent superpixels during the ROI pixel averaging (cf. section 2.3) using respectively 150 and 900 superpixels per frame. We can observe that the proposed method increased the estimation precision over the entire dataset for 7%. The averaged seeds value over adjacent superpixels performs even better with a 10% increased precision, which can be explained by the fact that the increased number of pixels considered for each superpixels values in our method, as long as the increased number of superpixels, provide the best compromise between segmentation quality and superpixel sizes. It is important to note here that we compare the proposed method to our previous implementation [3] because this latter had already been compared favorably with conventional supervised ROI selection techniques (such as face or skin detection) combined with conventional methods of RGB channel combinations such as CHROM or POS.

The computation time, which is up to several seconds for the TSP [5] implementation is reduced to 25 ms per

frame on single core implementation, which corresponds to 40 frames per second on this platform and with this level of optimizations.

4. Conclusion

Superpixels segmentation is a popular preprocessing step in many computer vision tasks. Unsupervised rPPG monitoring needs high quality segmentation as well as low computational burden. We have introduced in this paper, a low complexity superpixel algorithm capable of performing superpixels segmentation using only a fraction of the pixels of an image identifying iteratively and implicitly the boundaries. Experimental results on the BSDS500 dataset have demonstrated the efficiency of IBIS with a similar segmentation accuracy than SLIC but with a 8 times faster processing. At the same time, the quality of the segmentation and the temporal regularity of the propagation assure great results considering rPPG with the heart-rate estimation bringing real-time capabilities to unsupervised rPPG methods.

We can anticipate that the proposed method will not be very robust on videos presenting important movements. In future works, we will evaluate precisely this observation. Then, we will study further the block definition as well as the impact on the segmentation quality and the running-time of various strategies to select the evaluated pixels. We also plan to implement the algorithm on massively parallel platform such as GPU target to propose ultra-fast rPPG measurement. Hence rPPG could be part of a more complex real-time process.

Acknowledgments

This research was supported by the Conseil Régional de Bourgogne Franche-Comté, France and the Fond Européen de Développement Régional (FEDER).

References

- [1] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.
- [2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. Slic superpixels compared to state-of-the-art

- superpixel methods. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 34:2274–2282, 2012.
- [3] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 2017.
- [4] F. Bousefsaf, C. Maoui, and A. Pruski. Continuous wavelet filtering on webcam photoplethysmographic signals to remotely assess the instantaneous heart rate. *Biomedical Signal Processing and Control*, 8:568–574, 2013.
- [5] J. Chang, D. Wei, and J. Fisher. A video representation using temporal superpixels. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 2051–2058, 2013.
- [6] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(6):603–619, 2002.
- [7] S. Crommelinck, R. Bennett, M. Gerke, M. N. Koeva, M. Y. Yang, and et al. Isprs annals of the photogrammetry, remote sensing and spatial information sciences. *Gottingen, IV-2(W3):9–16*, 2017.
- [8] O. Csillik. Fast segmentation and classification of very high resolution remote sensing data using slic superpixels. *Remote Sensing*, 9(3):9–16, 2017.
- [9] G. de Haan and V. Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Trans. Biomed. Eng.*, 60:2878–2886, 2013.
- [10] M. V. den Bergh, X. Boix, G. Roig, and L. V. Gool. Seeds: Superpixels extracted via energy-driven sampling. *IEEE Conf. on Computer Vision and Pattern Recognition*, 13(6):603–619, 2013.
- [11] A. Guazzi, M. Villarroel, J. Jorge, J. Daly, M. Frise, P. Robbins, and L. Tarassenko. Non-contact measurement of oxygen saturation with an rgb camera. *Biomedical Optics Express*, 6:3320–3338, 2015.
- [12] G. D. Haan and A. V. Leest. Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiological measurement*, 35(9):1913, 2014.
- [13] C.-H. Huang, W.-A. Wang, S.-Y. Lin, and Y.-Y. Lin. Useq: Ultra-fast superpixel extraction via quantization. *23rd Intern. Conf. on Pattern Recognition*, 2016.
- [14] L. Jingjing, C. Ying, Z. Cheng, Y. Hua, and Z. Li. Tracking using superpixel features. *IEEE Conf. on Measuring Technology and Mechatronics Automation*, 2016.
- [15] A. Kamshilin, E. Nippolainen, I. Sidorov, P. Vasilev, N. Erofeev, N. Podolian, and R. Romashko. A new look at the essence of the imaging photoplethysmography. *Scientific Reports*, 5, 2015.
- [16] B. Kim and S. Yoo. Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE Trans. Biomed. Eng.*, 53:566–568, 2006.
- [17] S. Kwon, J. Kim, D. Lee, and K. Park. Roi analysis for remote photoplethysmography on facial video. In *Proc. EMBS*, pages 851–862, 2015.
- [18] M. Lewandowska, J. Rumiński, T. Kocejko, and J. Nowak. Measuring pulse rate with a webcam, a non-contact method for evaluating cardiac activity. In *Computer Science and Information Systems (FedCSIS), 2011 Federated Conf on*, pages 405–410. IEEE, 2011.
- [19] Z. Li, X.-M. Wu, and S.-F. Chang. Segmentation using superpixels: A bipartite graph partitioning approach. *IEEE Conf. on Computer Vision and Pattern Recognition*, 2012.
- [20] Z. Liu, X. Zhang, S. Luo, and O. L. Meur. Superpixel-based spatiotemporal saliency detection. *IEEE Trans. on Circuits and Systems for Video Technology*, pages 1522–1540, 2014.
- [21] R. Macwan, Y. Benezeth, and A. Mansouri. Remote photoplethysmography with constrained ica using periodicity and chrominance constraints. *Biomedical engineering online*, 17(1):22, 2018.
- [22] R. Macwan, Y. Benezeth, K. Nakamura, R. Gomez, Y. Wu, and A. Mansouri. Parameter-free adaptive step-size multi-objective optimization applied to remote photoplethysmography. *IEEE Intern. Conf. on Biomed. Health Inform.*, 2018.
- [23] G. Mori. Guiding model search using segmentation. *Intern. Conf. on Computer Vision*, 2:1417–1423, 2005.
- [24] M. Poh, D. McDuff, and R. Picard. Advancements in non-contact, multiparameter physiological measurements using a webcam. *IEEE Trans. Biomed. Eng.*, 58:7–11, 2011.
- [25] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- [26] Y. Sun, S. Hu, V. Azorin-Peris, S. Greenwald, J. Chambers, and Y. Zhu. Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. *Journal of Biomedical Optics*, 16, 2011.
- [27] Y. Sun and N. Thakor. Photoplethysmography revisited: From contact to noncontact, from point to imaging. *IEEE Trans. Biomed. Eng.*, 63:463–477, 2016.
- [28] H. Tasli, A. Gudi, and M. Uyl. Remote ppg based vital sign measurement using adaptive facial regions. *IEEE Intern. Conf. on Image Processing*, pages 1410–1414, 2014.
- [29] X. Teng and Y. Zhang. The effect of contacting force on photoplethysmographic signals. *Physiological Measurement*, 25:1323–1335, 2004.
- [30] W. Verkrusse, L. O. Svaasand, and J. S. Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16:21434–21445, 2008.
- [31] W. Wang, S. Stuijk, and G. de Haan. Exploiting spatial-redundancy of image sensor for motion robust rppg. *IEEE Trans. Biomed. Eng.*, 62:415–425, 2015.
- [32] W. Wang, S. Stuijk, and G. de Haan. Unsupervised subject detection via remote ppg. *IEEE Trans. Biomed. Eng.*, 62:2629–2637, 2015.
- [33] W. Wang, S. Stuijk, and G. de Haan. Novel algorithm for remote photoplethysmography: Spatial subspace rotation. *IEEE Trans. Biomed. Eng.*, 63:1974–1984, 2016.
- [34] W. Wang, S. Stuijk, and G. de Haan. Living-skin classification via remote-ppg. *IEEE Trans. Biomed. Eng.*, 64(12):2781–2792, 2017.
- [35] N. Xuesong, H. Hu, S. Shiguang, and C. Xilin. Continuous heart rate measurement from face: A robust rppg approach with distribution learning. 2017.