

Detecção automática de pulso cardíaco utilizando imagens de vídeo

Gustavo L. Sandri, Ricardo L. de Queiroz e Eduardo Peixoto

Resumo— O fluxo sanguíneo através da pele de um indivíduo altera periodicamente a maneira como esta irradia luz. Esta pulsação pode ser captada por uma câmera de vídeo, permitindo que, através do processamento das imagens adquiridas, se determine a frequência cardíaca do indivíduo filmado. A medição remota do pulso cardíaco oferece maior conforto por evitar o uso de eletrodos. Neste artigo introduzimos uma nova metodologia para a detecção de pulso cardíaco reduzindo os efeitos do ruído. A técnica é aplicada à vídeos em cores do rosto, usando detecção de face e filtragem adaptativa para redução de ruído.

Palavras-Chave— Detecção de pulso, Frequência cardíaca, Fotopletismografia, Processamento de vídeo.

Abstract— The blood flow through the skin of an individual changes the way that light is irradiated. This pulse can be captured by a video camera allowing, by processing the acquired images, to determine the heart rate of the subject. The remote measurement of cardiac pulse provides more comfort because it avoids the use of electrodes. In this paper we introduce a new approach to heart rate detection reducing noise effects. The technique is applied to colored videos of the face, using face detection and adaptive filtering for noise reduction.

Keywords— Pulse detection, Heart Rate, Photoplethysmography, Video processing.

I. INTRODUÇÃO

A medida do pulso cardíaco é um sinal crítico para o diagnóstico médico. Entretanto, os métodos atuais utilizam equipamentos específicos e muitos exigem contato com a pele. Existe um interesse crescente em extrair este valor sem a necessidade de contato, aumentando o conforto para o sujeito, utilizando a fotopletismografia (do inglês *photoplethysmography*, PPG), que é uma técnica bem conhecida por capturar visualmente a perfusão do fluxo sanguíneo e detectar o ciclo cardíaco. Mais recentemente, tem sido demonstrado que o pulso cardíaco pode ser encontrado de forma eficiente utilizando câmeras comuns [7], [9], [11], [13], [15], permitindo assim que seja determinado remotamente por equipamentos de baixo custo, além de tornar possível o monitoramento de um indivíduo por longos períodos.

Wu *et. al.* [16] empregam *eulerian video magnification* para tornar visível, em um vídeo, as sutis variações da tonalidade da pele causado pela circulação do sangue. Entretanto, não tentam resolver o problema de detectar a frequência da pulsação.

Poh *et. al.* [10] e Pal *et. al.* [8] exploraram a possibilidade de medir o pulso cardíaco através de vídeos da face. Um retângulo contendo o rosto é definido como região de interesse, encontrado através do detector de face ‘Viola-Jones’ [14], e utilizam a média do valor dos pixels dentro desta região. Para

melhorar a qualidade da relação sinal/ruído eles empregam a separação cega de fontes através da análise de componentes independentes (do inglês *independent component analysis*, ICA) de forma a tentar isolar o sinal PPG contido nos três canais de cores. O sinal resultante é então transferido para o domínio de Fourier para se encontrar a frequência de máxima energia dentro da faixa de 40 a 240 batimentos por minuto, que corresponde às frequências esperadas para o pulso cardíaco.

Balakrishnan *et. al.* [2] propuseram um método que detecta a pulsação cardíaca a partir de micromovimentos do rosto causados pela circulação cardiovascular. Aplicam a análise de componentes principais para melhorar a robustez em relação ao ruído. A técnica permite que a detecção seja feita mesmo quando a pele não está visível, mas como se apoia no movimento da cabeça para fazer suas estimativas, exige que os voluntários estejam estáticos.

Li *et. al.* [5] utilizam *discriminative response map fitting* [1] para determinar de forma mais precisa a região de interesse no rosto e consideram somente os sinais provenientes do canal verde. Reduzem os artefatos de movimento rastreando o movimento da cabeça em pontos específicos através do algoritmo de Kanade-Lucas-Tomase [12]. Utilizam ainda o nível médio de verde no plano de fundo da cena para corrigir as variações de iluminação. Os sinais obtidos são divididos em segmentos e aqueles que possam ter sofrido a interferência de movimentos não rígidos são desconsiderados.

Neste artigo nós propomos um novo método para detecção remota de pulso cardíaco, baseado nos trabalhos de Poh *et. al.* [10] e Pal *et. al.* [8], com o objetivo de ser um algoritmo de baixa complexidade computacional, para permitir uma fácil implementação em dispositivos de baixo desempenho, e com maior robustez ao ruído. Os resultados obtidos com nossa técnica são comparados com aqueles obtidos através do método de Poh e verificamos uma substancial melhora da performance.

II. MÉTODO PROPOSTO

A Fig. 1 apresenta um diagrama de blocos do processamento utilizado para se obter a pulsação através do vídeo. $I[i]$ representa a imagem do i -ésimo quadro, que são adquiridos à uma taxa de amostragem constante, e $p[j]$ a j -ésima frequência da pulsação detectada. As pulsações são calculadas a cada meio segundo e necessitam de uma sequência de quadros para serem estimadas.

O processamento do sinal é similar àquele aplicado ao método de Poh, com exceção dos blocos destacados que, ou foram modificados significativamente, ou adicionados. Os blocos são explicados detalhadamente nas subseções a seguir.

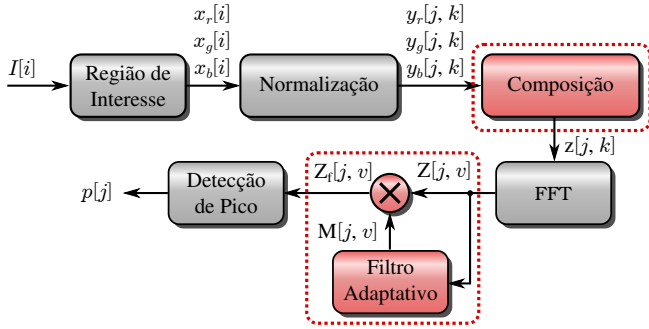


Fig. 1. Esquemático do processamento do sinal. Os sinais que acompanham as conexões entre os blocos representam as informações trocadas.

A. Região de Interesse

Nosso algoritmo utiliza a face do indivíduo para fazer a detecção por ser uma região fácil de ser encontrada por algoritmos já conhecidos da literatura e por apresentar uma significativa variação de tonalidade com a pulsação, capaz de ser captada por uma câmera.

Nós utilizamos o algoritmo de detecção de face do OpenCV, baseado nos trabalhos de Viola e Jones [14] e de Lienhart e Maydt [6] para encontrar as coordenadas da face em cada quadro do vídeo. Este algoritmo utiliza um classificador em cascata que usa 14 características do tipo Haar da imagem, treinado com exemplos positivos e negativos de faces. No caso de nosso trabalho, utilizamos o classificador de face frontal, já treinado, disponível com o OpenCV 2.0.

Para cada região detectada como face, o algoritmo retorna as coordenadas x e y e a largura e altura que determinam um retângulo em torno da região encontrada. A região de interesse foi escolhida como sendo o retângulo de mesma altura e 60% da largura da região encontrada pelo algoritmo, centrada no meio da face (ver Fig. 2).

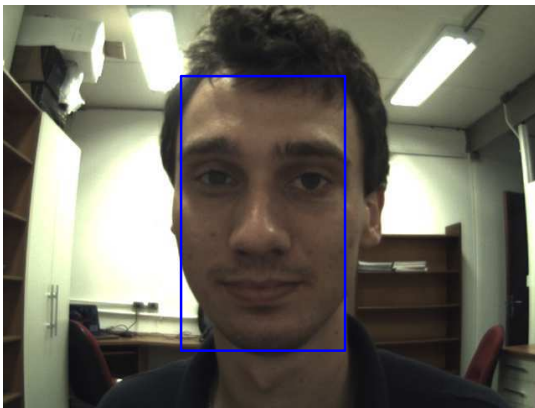


Fig. 2. Região de interesse

Caso nenhuma face seja encontrada no quadro atual, as coordenadas do quadro anterior são utilizadas.

Os sinais $x_r[i]$, $x_g[i]$ e $x_b[i]$ correspondem à média dos valores dos pixels para os canais vermelho, verde e azul, respectivamente, calculados dentro da região de interesse da imagem $I[i]$.

B. Normalização

A normalização consiste em remover a média e ajustar a amplitude dos sinais $x_c[i]$ (onde c corresponde a um dos canais de cor) de tal forma que este tenha variância unitária, resultando no sinal

$$y_c[j, k] = \frac{x_c[Nj + k] - \mu_c[j]}{\sigma_c[j]}, \quad k \in \Theta, \quad (1)$$

onde a média e a variância (μ e σ , respectivamente) são dados por:

$$\mu_c[j] = \frac{1}{M} \sum_{k \in \Theta} x_c[Nj + k] \quad (2)$$

e

$$\sigma_c[j]^2 = \frac{1}{M} \sum_{k \in \Theta} (x_c[Nj + k] - \mu_c[j])^2. \quad (3)$$

O número de quadros contidos em 30 segundos, dado por M , define o tamanho da janela Θ dentro da qual os sinais são observados, tal que

$$\Theta = \left\{ k \in \mathbb{Z} \left| -\frac{M}{2} \leq k < \frac{M}{2} \right. \right\}. \quad (4)$$

N é o passo utilizado pelo algoritmo e é dado pela quantidade de quadros contidos em 0.5 segundos.

Para um j fixo, $y_c[j, k]$ é um vetor de duração 30 segundos sobre o qual determinamos a pulsação do indivíduo. Conforme o valor de j progride, a janela vai se deslocando de meio segundo ao longo do sinal, com uma sobreposição de 96,67% (29/30).

C. Composição

O sinal de interesse em nosso estudo é a onda provocada pelo fluxo sanguíneo. O sangue possui grande quantidade de células vermelhas responsáveis pelo transporte de oxigênio. Estas células contêm hemoglobina, que absorvem a luz de forma diferente, dependendo do fato de estarem ou não ligadas ao oxigênio ou monóxido de carbono. Duas destas variações são de interesse neste trabalho: a oxi-hemoglobina e a desoxi-hemoglobina.

A oxi-hemoglobina é formada durante a respiração, quando o oxigênio se liga ao grupo hemo da proteína nas células vermelhas do sangue. Quando esta não está ligado ao oxigênio, é denominada desoxi-hemoglobina. A Fig. 3 apresenta o espectro de absorção para a hemoglobina.

Assim, a medida que o ciclo evolui, a densidade de oxi-hemoglobina e desoxi-hemoglobina na pele vai variando periodicamente. A quantidade de células vermelhas na proximidade da pele também muda, o que altera o comprimento médio do caminho que a luz deve percorrer antes de ser refletida. Tudo isso afeta a maneira como a pele emite a radiação do ambiente, que é captada pela câmera.

A câmera divide a luz de entrada em três canais distintos através de um filtro de cor. Cada canal receberá uma versão atenuada do sinal do fluxo sanguíneo, acrescido de ruídos provenientes de artefatos de movimento, qualidade dos

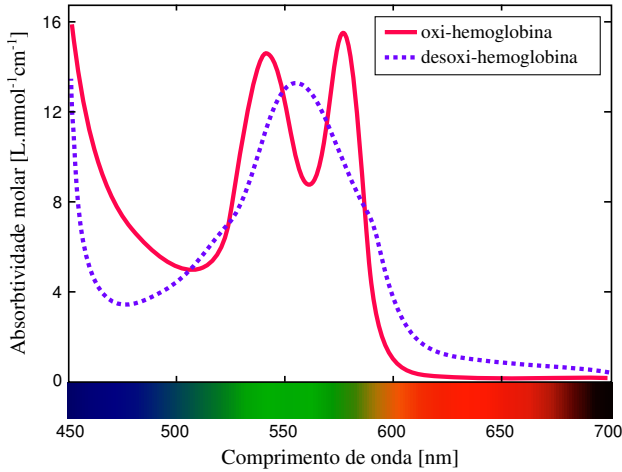


Fig. 3. Absorção de luz pela hemoglobina de um adulto saudável no espectro da luz visível. Adaptado de Zijlstra *et. al.* [17]

sensores, entre outros. Desta forma, os sinais $y_c[j, k]$ podem ser interpretados como sendo compostos pelo sinal da onda do fluxo sanguíneo multiplicado por uma constante intrínseca do comprimento de onda do filtro de cor do canal somada à um ruído proveniente de artefatos de movimento, variação na iluminação, erros na determinação da região de interesse, entre outros.

Podemos combinar linearmente cada um desses canais em um único vetor de tal forma a maximizar a energia do sinal de interesse em relação ao ruído, resultando em

$$z[j, k] = \alpha_r y_r[j, k] + \alpha_g y_g[j, k] + \alpha_b y_b[j, k]. \quad (5)$$

Nos trabalhos de Poh *et. al.* [10] e Pal *et. al.* [8], os valores de α_r , α_g e α_b são determinados por ICA para cada instante j . Estes valores variam de um vídeo para outro e ao longo do tempo.

Entretanto, observamos em nosso trabalho que a razão entre os valores α é relativamente constante, sendo pouco dependente da tonalidade da pele do indivíduo e da iluminação da cena. A Fig. 4 apresenta a função distribuição de probabilidade de α assumir um dado valor, estimada a partir do histograma dos valores encontrados em 5 vídeos testados, utilizando ICA, após multiplicar por uma constante de forma a satisfazer as condições $\alpha_g > 0$ e $|\alpha_r| + |\alpha_g| + |\alpha_b| = 1$.

Como estes valores tendem a se concentrar em uma dada região, assumimos eles como constantes dadas pelo valor esperado, calculados a partir da distribuição de cada uma. Os valores utilizados foram $\alpha_r = -0.184$, $\alpha_g = +0.501$ e $\alpha_b = -0.315$. Os valores esperados calculados estão condizentes com a literatura, pois o canal verde é o que contém a maior energia para o sinal fotopletiomográfico entre os três canais, seguido pelo canal vermelho [5], [13]. Assim, a combinação dos canais vermelho e azul com os pesos α_r e α_b é composta principalmente de ruído e este ruído é subtraído do canal verde.

D. FFT e Filtro Adaptativo

Para um dado j fixo, o sinal $z[j, k]$ é preenchido com zeros para que tenha um total de 2^{14} elementos. É calculado

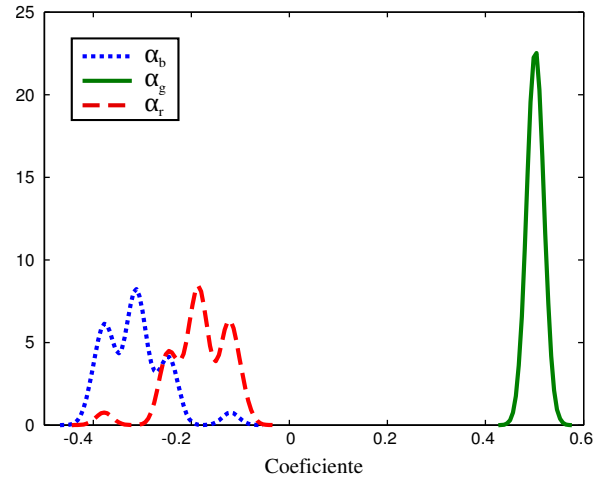


Fig. 4. Função distribuição de probabilidade estimada para os valores encontrados para as constantes α utilizando a Análise de Componentes Independentes

então o valor absoluto da transformada rápida de Fourier (do inglês *fast Fourier transform*, FFT) deste sinal estendido, resultando em $Z[j, v]$, onde v é o índice do componente de frequência. Isolamos os coeficientes que correspondam à frequência entre 30 e 240 batimentos por minuto (bpm), que é a faixa onde esperamos encontrar a pulsação do sujeito. Ou seja, $2^{14} \cdot 30 / f_s \leq v \leq 2^{14} \cdot 240 / f_s$, onde f_s é a taxa de amostragem da câmera, em quadros por minuto.

É então aplicado uma filtragem adaptativa em $Z[j, v]$, multiplicando este sinal por uma máscara $M[j, v]$, resultando no sinal

$$Z_f[j, v] = Z[j, v]M[j, v]. \quad (6)$$

Esta máscara é determinada supondo que a pulsação cardíaca varia pouco no período de meio segundo. Portanto o pico atual deve estar próximo daquele anteriormente encontrado. A máscara é calculada convoluindo a transformada dos dois sinais precedentes por um filtro passa-baixa $h[v]$ de modo a dilatar a base dos picos para acomodar melhor pequenas variações de frequência. Neste trabalho, $h[v]$ é uma função triangular centrada em torno da frequência zero, variando entre -2 e 2 bpm, tal como na Figura 5. Assim, a máscara é dada por:

$$M[j, v] = (h[v] * Z[j-1, v])(h[v] * h[v] * Z[j-2, v]), \quad (7)$$

onde $*$ denota convolução.

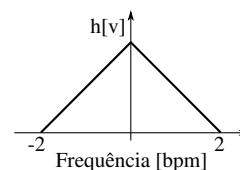


Fig. 5. Filtro passa-baixa empregado para calcular a máscara do filtro adaptativo

O objetivo desta máscara é atenuar as regiões onde não esperamos encontrar o pico da pulsação cardíaca, reduzindo o ruído do sinal, tal como representado na Fig. 6.

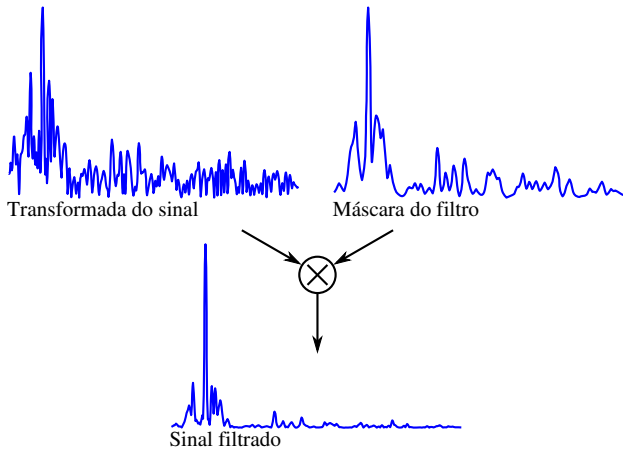


Fig. 6. Filtragem adaptativa: A transformada do sinal é multiplicado por uma máscara de forma a atenuar regiões onde não esperamos encontrar um pico de batimento cardíaco

Como as componentes do sinal provenientes de ruído costumam variar de frequência ao longo do tempo, ao multiplicarmos a transformada dos dois últimos sinais acabamos dando menor ênfase à estas posições da frequência.

Finalmente, a partir do sinal filtrado, é encontrada a posição do pico de maior amplitude e sua frequência é atribuída à pulsação cardíaca do sujeito sendo monitorado. Caso a diferença entre a frequência encontrada e a anterior seja superior à 12 bpm , o algoritmo procura entre os 4 maiores picos qual deles é mais próximo da frequência anterior, possuindo uma diferença inferior a 12 bpm . Se nenhum dos 4 picos atender esta exigência, a frequência do maior pico é utilizada.

III. RESULTADOS E DISCUSSÃO

Para testar nosso algoritmo utilizamos cinco vídeos capturados de cinco voluntários (um vídeo para cada voluntário), sendo 4 do sexo masculino e 1 do sexo feminino, com idades entre 20 e 30 anos, de tonalidade de pele distintas. Os vídeos tem duração de 1 minuto. Os voluntários ficaram sentados em frente à uma câmera e foram aconselhados a se mover o mínimo possível. Os vídeos foram capturados com a câmera Firefly MV FMVU-03MTC da Point Grey, que possui uma resolução de 0,3 megapixels (640x480) e taxa de amostragem de 60 quadros por segundo. Cada quadro capturado é enviado sem compressão diretamente para um computador por uma conexão usb e são armazenados em um arquivo RAW. Os quadros capturados apresentam um padrão de Bayer [3] do tipo 'RGGB' com 8 bits por pixel. As imagens $I[i]$ foram obtidas a partir do arquivo RAW salvo na memória do computador, convertendo a matriz com o padrão de Bayer em uma imagem RGB.

Um oxímetro de dedo foi utilizado para medir a pulsação cardíaca ao mesmo tempo que o vídeo era adquirido. O oxímetro foi filmado por uma câmera comum, diferente da usada para o estudo, para que a leitura de sua medida pudesse

ser feita após a captura e sincronizado com o vídeo. O oxímetro possui uma precisão de leitura de 2 batimentos por minuto.

Os resultados obtidos foram comparados com a nossa implementação do algoritmo de Poh. ICA foi executada utilizando a implementação de Cardoso [4].

As Fig. 7 e 8 apresentam os resultados obtidos ao longo do tempo para dois dos vídeos. Ambos algoritmos apresentam resultados parecidos nas regiões em que a relação sinal ruído é relativamente boa, mas a medida que esta relação vai se degradando nosso algoritmo passa a apresentar um melhor desempenho, sendo mais robusto ao ruído, como pode ser visto na Fig. 8. Isto é devido principalmente à filtragem adaptativa, que consegue eliminar boa parte do ruído, como mostrado na Fig. 9, que apresenta a transformada do sinal utilizado para determinar a pulsação cardíaca para o vídeo 2, usando o método de Poh e o nosso.

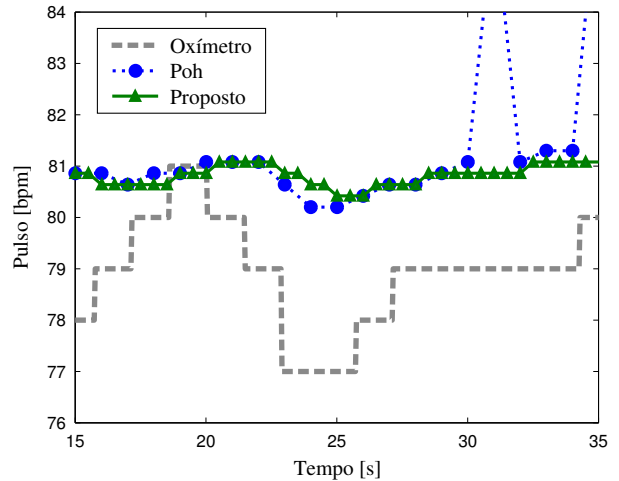


Fig. 7. Detecção da pulsação ao longo do tempo para o vídeo 5

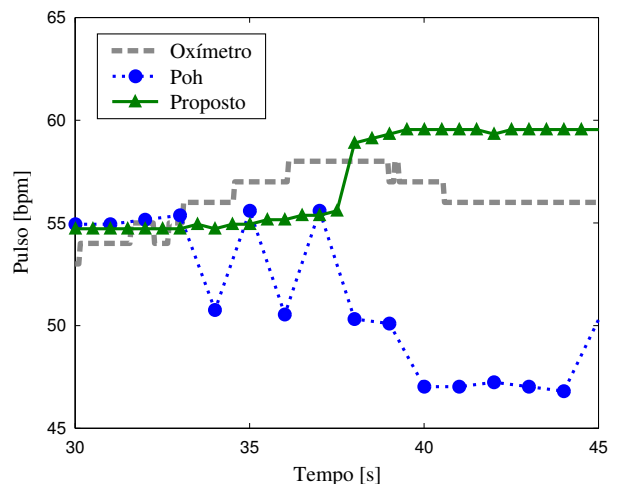


Fig. 8. Detecção da pulsação ao longo do tempo para o vídeo 2

A Tabela I resume os resultados obtidos para os cinco vídeos. Para facilitar a leitura é apresentado o valor médio encontrado durante um período de 6 segundos com ambos algoritmos.

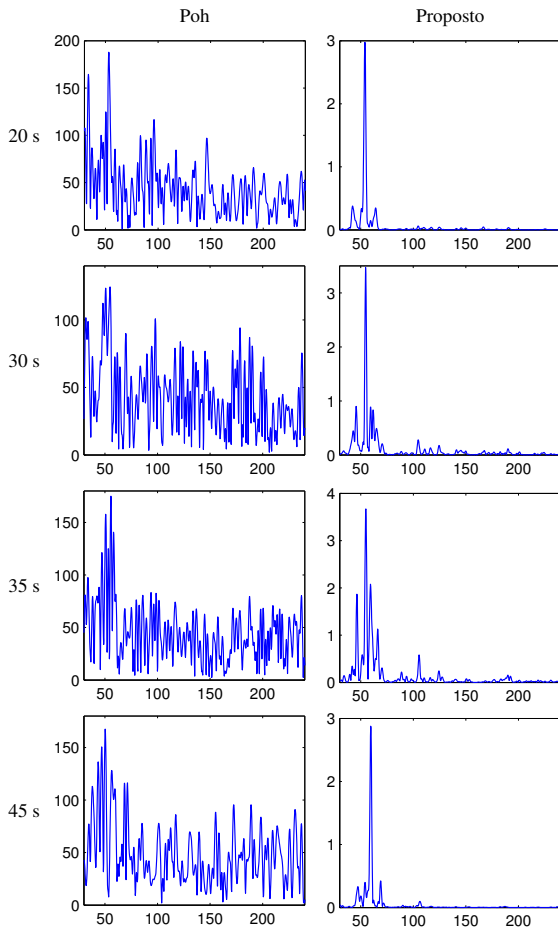


Fig. 9. Sinal transformado no domínio da frequência utilizado para a detecção do pulso cardíaco, utilizando os métodos de Poh e o proposto, para o vídeo 2, nos instantes de tempo de 20, 30, 35 e 45 segundos

IV. CONCLUSÃO

Neste trabalho foi descrito, implementado e avaliado uma nova metodologia para estimar o pulso cardíaco através de vídeos gravados da face humana usando uma câmera em iluminação ambiente. Comparamos nosso resultado com aquele de Poh e observamos uma substancial melhora em relação à robustez ao ruído. Além disso, o algoritmo proposto não utiliza ICA, que possui um custo computacional elevado, facilitando a aplicação em dispositivos de baixa performance computacional.

É previsto para trabalhos futuros utilizar eletrocardiograma para se obter de forma mais precisa a frequência do batimento cardíaco empregado para avaliar a performance da técnica e utilizar detecção de pele e movimento para reduzir artefatos.

REFERÊNCIAS

- [1] A. Asthana, S. Zafeiriou, Shiyang Cheng, and M. Pantic. Robust discriminative response map fitting with constrained local models. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3444–3451, June 2013.
- [2] G. Balakrishnan, F. Durand, and J. Guttag. Detecting pulse from head motions in video. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 3430–3437, June 2013.
- [3] B.E. Bayer. Color imaging array, July 20 1976. US Patent 3,971,065.
- [4] Jean-François Cardoso. High-order contrasts for independent component analysis. *Neural Comput.*, 11(1):157–192, January 1999.

TABELA I

MÉDIA DA PULSAÇÃO CARDÍACA OBTIDOS PELOS ALGORITMOS DENTRO DE UMA JANELA DE 6 SEGUNDOS PARA OS CINCO VÍDEOS TESTADOS.

Intervalo [segundos]	PULSAÇÃO [BPM]			ERRO ABSOLUTO		
	Oxímetro	Poh	Proposto	Poh	Proposto	
1	18 – 24	65.00	68.68	68.77	3.68	3.77
	25 – 31	65.29	68.77	68.77	3.49	3.49
	32 – 38	65.43	69.28	69.25	3.85	3.82
	39 – 45	66.14	69.40	69.37	3.26	3.23
2	18 – 24	53.71	36.82	54.15	16.89	0.43
	25 – 31	53.57	54.12	54.65	0.54	1.08
	32 – 38	56.57	53.33	55.50	3.24	1.07
	39 – 45	56.29	47.93	59.48	8.35	3.20
3	18 – 24	75.00	74.36	74.77	0.64	0.23
	25 – 31	75.00	71.35	70.69	3.65	4.31
	32 – 38	75.00	70.28	70.28	4.72	4.72
	39 – 45	79.00	70.12	70.16	8.88	8.84
4	18 – 24	72.43	79.23	79.20	6.80	6.77
	25 – 31	74.86	77.75	78.35	2.89	3.49
	32 – 38	76.43	77.78	77.34	1.35	0.92
	39 – 45	79.14	79.95	76.81	0.81	2.33
5	18 – 24	79.29	80.83	80.26	1.54	1.57
	25 – 31	78.43	81.39	80.67	2.96	2.24
	32 – 38	80.00	84.19	81.83	4.19	1.83
	39 – 45	81.57	75.46	85.82	6.11	4.25

- [5] Xiaobai Li, Jie Chen, Guoying Zhao, and Matti Pietikainen. Remote heart rate measurement from face videos under realistic situations. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.
- [6] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I-900–I-903 vol.1, 2002.
- [7] Yuka Maeda, Masaki Sekine, and Toshiyo Tamura. The advantages of wearable green reflected photoplethysmography. *Journal of Medical Systems*, 35(5):829–834, 2011.
- [8] Arpan Pal, Aniruddha Sinha, Anirban Dutta Choudhury, Tanushyam Chattopadhyay, and Aishwarya Viswanathan. A robust heart rate detection using smart-phone video. In *Proceedings of the 3rd ACM MobiHoc Workshop on Pervasive Wireless Healthcare, MobileHealth '13*, pages 43–48, New York, NY, USA, 2013. ACM.
- [9] P. Pelegris, K. Banitsas, T. Orbach, and K. Marias. A novel method to detect heart beat rate using a mobile phone. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 5488–5491, Aug 2010.
- [10] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010.
- [11] T Pursche, J Krajewski, and R Moeller. Video-based heart rate measurement from human faces. *Electronics (ICCE)*, pages 544–545, 2012.
- [12] Carlo Tomasi and Takeo Kanade. Detection and tracking of point features. Technical report, International Journal of Computer Vision, 1991.
- [13] Wim Verkruijsse, Lars O. Svaasand, and J. S. Nelson. Remote plethysmographic imaging using ambient light. *Opt. Express*, 16(26):21434–21445, Dec 2008.
- [14] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I-511–I-518 vol.1, 2001.
- [15] F. P. Wieringa, F. Mastik, and Steen. Contactless multiple wavelength photoplethysmographic imaging: A first step toward 'spo2 camera' technology. *Annals of Biomedical Engineering*, 33(8):1034–1041, 2005.
- [16] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William T. Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph. (Proceedings SIGGRAPH 2012)*, 31(4):651–658, 2012.
- [17] W. G. Zijlstra, A. Buursma, and W. P. Meeuwse-van der Roest. Absorption spectra of human fetal and adult oxyhemoglobin, deoxyhemoglobin, carboxyhemoglobin, and methemoglobin. *Clinical Chemistry*, 37(9):1633 – 1638, 1991.