

Estimação de Pose para a Fisioterapia

Gabriel Luís de Araújo Freitas (15/0008872)

Faculdade do Gama - FGA
Universidade de Brasília - UnB
Brasília, Brasil
Email: gabrieluis10@gmail.com

Victor Oliveira Corrieri de Macedo (14/0164961)

Faculdade do Gama - FGA
Universidade de Brasília - UnB
Brasília, Brasil
Email: victormacedo10@yahoo.com.br

Resumo—Clínicas de fisioterapia e reabilitação têm dificuldades de custo e praticidade em acompanhamentos contínuos e tendem a estender o tratamento para fora do consultório. A realização de exercícios em casa limita a supervisão e correção dos exercícios, prejudicando a recuperação do paciente. Este trabalho trata da proposta de um aparelho de auxílio em exercícios de fisioterapia, por meio de um algoritmo de estimação de pose embarcado em um *system on chip* (SoC). O algoritmo, baseado em redes neurais convolucionais, usa um vídeo de referência para dar um *feedback* visual em tempo real ao usuário. Assim, permitindo a execução de um exercício mais próximo ao proposto, além do acompanhamento à distância pelo fisioterapeuta.

I. INTRODUÇÃO

Em vista de uma tendência de automatização e personalização, sistemas assistivos vêm ganhando bastante espaço nas áreas da saúde. A geração de informações e registros temporais facilitam e embasam tomadas de decisão e diagnósticos médicos. Tal abordagem traz para áreas, como a fisioterapia, uma ferramenta para melhor acompanhamento do paciente, especialmente para tratamentos em casa.

Recentes avanços tecnológicos em visão computacional possibilitaram o monitoramento de atividades físicas com baixo custo e alto desempenho. A área de captura de movimento teve um crescente interesse de estudo, sobretudo devido a popularização de produtos como o Microsoft Kinect. Trabalhos que obtiveram resultados no estado da arte em estimação de pose disponibilizaram seus algoritmos para uso público [1], impulsionando as mais diversas aplicações [2]. Além disso, grandes bases de dados foram criadas permitindo testes e otimizações de novos algoritmos [3].

Diversos trabalhos foram desenvolvidos no intuito de monitorar e auxiliar o tratamento fisioterápico doméstico [4]. Porém a maioria faz uso de marcadores ou sensores acoplados ao paciente, o que dificulta a praticidade e aceitação do dispositivo. Outros artigos propuseram o uso do Microsoft Kinect para atacar o problema [5], contudo, apresentam um maior custo de hardware e se restringem ao grupo de usuários que já possuem a tecnologia. Neste sentido, a implementação de um sistema embarcado de baixo custo com uma câmera simples acoplada, mostra-se como abordagem viável para o problema.

II. JUSTIFICATIVA E OBJETIVOS

Muitos dos tratamentos de fisioterapia necessitam de que o paciente repita em casa os exercícios realizados em consul-

tório. Porém, sem o acompanhamento adequado o tratamento pode perder efetividade e se estender por mais tempo. Logo, ferramentas que ajudem a realização do movimento e que possibilitem seu monitoramento são úteis no contexto da fisioterapia.

Dada a oportunidade de desenvolvimento de tecnologias que auxiliem a execução de exercícios fisioterápicos, este projeto visa apresentar uma solução, utilizando como plataforma a *Raspberry Pi 3* e algoritmos de estimação de pose, para uso em contexto doméstico. O produto, programável pelo fisioterapeuta, fornece ao paciente um *feedback* visual para a reprodução dos movimentos corretos de maneira mais adequada.

Juntamente com o auxílio na execução dos exercícios, são produzidos dados quantitativos referentes à qualidade dos movimentos que, fornecidos ao fisioterapeuta, são utilizados para o aprimoramento do tratamento. O equipamento, portanto, é uma ferramenta para facilitar a realização dos movimentos e possibilitar um acompanhamento mais efetivo dos pacientes.

III. FUNCIONAMENTO PROPOSTO

O sistema funciona a partir de um vídeo de referência com o movimento certo realizado pelo fisioterapeuta. Uma vez filmado, o vídeo é previamente processado para extrair os pontos de articulação e determinar suas conexões, de forma a criar uma representação em *stickman*. O paciente então posiciona o dispositivo com a câmera apontada para o plano de maior informação do exercício, conforme determinado pelo fisioterapeuta, e seu vídeo é apresentado em um monitor de sua escolha. Para adequar o *stickman* ao paciente, uma etapa de calibração será necessária. Nesta etapa, as proporções do usuário são extraídas pelo mesmo algoritmo de estimação de pose e aplica-se uma redimensionalização do *stickman* salvo. Durante o exercício, o *stickman* é sobreposto no vídeo e o paciente tem como objetivo imitar seu movimento.

Deseja-se que os exercícios a serem feitos possam ser armazenados no sistema pelo fisioterapeuta. O paciente ao ligar o equipamento em uma televisão poderá escolher, por meio de um controle *bluetooth*, qual exercício realizar, terá acesso também às mensagens enviadas pelo fisioterapeuta e a um menu de calibração. Uma vez escolhido o exercício, ele será iniciado após um segundo comando dado pelo controle, para que o paciente posicione-se corretamente. Por fim, poderão ser feitas pausas no movimento e o usuário poderá retornar

ao menu de seleção. Dado o funcionamento desejado foram levantados os seguintes requisitos para o seu projeto:

- Dispositivo que exiba as imagens em uma televisão;
- Facilidade de configuração pelo fisioterapeuta;
- Uso de alguma metodologia para detecção de pose;
- Comunicação serial via *bluetooth* para o controle ;
- Controle ergonômico para não atrapalhar a execução do movimento;
- Armazenamento e apresentação gráfica dos dados obtidos;
- Canal de comunicação paciente/fisioterapeuta via internet;
- Baixo custo;
- Interface agradável ao usuário.

IV. ARQUITETURA PROPOSTA

Para apresentar a viabilidade de implementação do sistema embarcado na *Raspberry Pi 3* serão mostrados a seguir o funcionamento de módulos fundamentais na composição do sistema como um todo.

A. Extração do movimento do vídeo

O algoritmo de detecção de pose implementado é baseado em uma rede neural convolutiva (CNN), treinada em um amplo banco de imagens para captura de movimentos [3]. A arquitetura da rede foi desenvolvida em [1], sendo constituída por duas partes, detalhadas na figura 1:

- Classificação de regiões:** Estima mapas de probabilidade que indicam as regiões de maior probabilidade da localização de cada ponto de articulação. Utiliza-se então uma política *greedy* para extrair um único ponto para cada mapa. Essa parte da rede é composta de uma CNN com suas 10 camadas iniciais adquiridas da arquitetura VGGNet, por transferência de aprendizado, e múltiplos estágios ramificados um para cada parte do corpo.
- Afinidade entre partes:** Após a identificação de pontos individuais, a próxima etapa necessária é definir a conexão entre eles. Para isso, são gerados mapas de confiança e afinidade com o intuito de estimar regiões de probabilidade para a ligação entre pontos adjacentes. Novamente, uma inferência *greedy* é aplicada, definindo conexões e classificando cada região. A saída produzida pela rede é um vetor de 18 componentes, compondo as articulações principais e os pontos faciais, porém, apenas 13 componentes foram utilizados para o projeto, como ilustrado na figura 2



Figura 2. Pontos de articulação resultantes da estimação de pose.

A etapa de aquisição dos pontos no vídeo de referência foi realizada externamente em uma CPU, devido as limitações de poder computacional da placa utilizada. Para isso, o algoritmo de estimação de pose, descrito na seção anterior, foi executado *frame a frame* no vídeo e os pontos obtidos foram salvos em um arquivo texto, que foi armazenado na *Raspberry*.

B. Calibração do stickman

As proporções corporais variam de acordo com o indivíduo, este fato deve ser considerado no desenvolvimento do projeto. Para que o paciente possa seguir o movimento desejado de maneira mais natural é importante que as dimensões do *stickman* sejam coerentes com as suas próprias. Assim o *stickman* que executa o movimento de referência deve ser redimensionado para que as distâncias entre as articulações estejam próximas das do paciente. O módulo chamado de calibração é responsável por esta tarefa.

A etapa de calibração tem como entrada dois conjuntos de pontos, denominados “professor” e “aluno”. Cada um deles é a saída do algoritmo de detecção de pose para uma imagem em que a pessoa que executa o movimento está em uma posição padrão que favorece a visualização das articulações a serem vistas pela câmera durante a execução do exercício. O conjunto “professor” refere-se à pessoa que executou o movimento gravado, que será imitado pelo paciente, correspondente ao conjunto “aluno”.

Em cada um dos *frames* do vídeo deve-se manter o ângulo formado entre a horizontal e a reta que une dois pontos de articulação do conjunto “professor”, mas a distância entre os mesmos dois pontos deve ser a observada nos pontos correspondentes no conjunto “aluno”. Este procedimento é

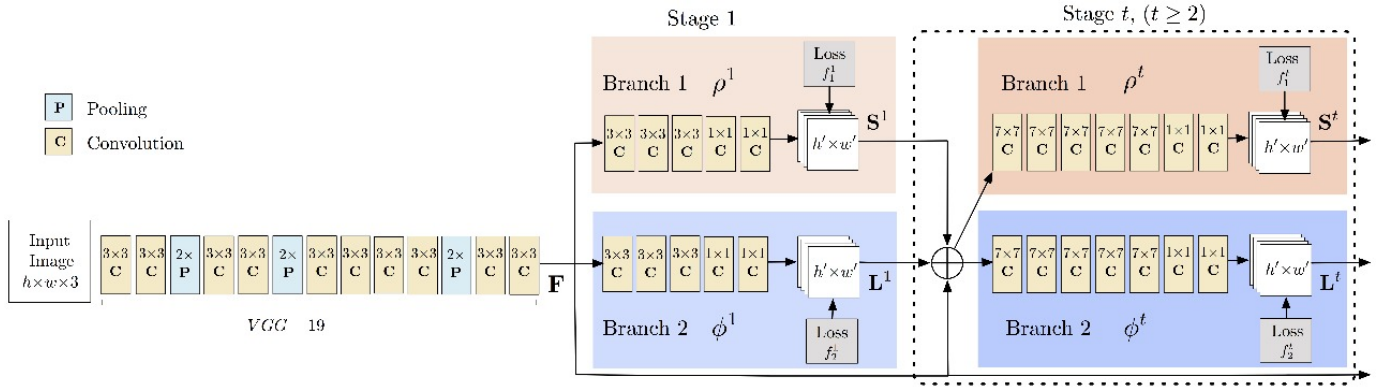


Figura 1. Arquitetura CNN para detecção de pose. Fonte: [1].

feito percorrendo todos os pares conectados da imagem de forma que cada ponto é movimentado uma única vez. As novas coordenadas do ponto $[x_m \ y_m]^T$ movimentado a cada iteração é dado por:

$$[x_m \ y_m]^T = [x_p \ y_p]^T + d \cdot [\cos(\theta) \ \sin(\theta)]^T + \mu \quad (1)$$

Em que d é a distância vista no conjunto “aluno”, θ o ângulo visto no conjunto “professor”, $[x_p \ y_p]^T$ são as coordenadas do ponto que se mantém fixo na iteração e μ é um vetor com duas posições, atualizado a cada nova iteração, que é utilizado para fazer correções relativas ao movimento global do *stickman*.

A figura 3(a) apresenta o resultado desse algoritmo em que o *stickman* em amarelo é o resultado do algoritmo propriamente dito, o *stickman* em preto é a entrada do algoritmo como “aluno” e os pontos vistos na figura 3(b) também são dados como entradas, mas para o conjunto “professor”.



(a) Saída do algoritmo e pontos do “aluno” (b) Pontos do “professor”

Figura 3. Imagens de teste para o algoritmo de calibração

C. Sobreposição do Vídeo

Uma vez calibrado, o usuário poderá escolher um exercício para executar dentre os gravados na memória. Nesta etapa, a

câmera acoplada ao sistema é acionada e os pontos extraídos para o exercício são exibidos sobrepostos às imagens captadas em tempo real. O usuário tem controle sobre o movimento realizado pelo *stickman* sobreposto, podendo pausar e recomeçar a execução para um melhor acompanhamento do exercício proposto. Além disso, o usuário pode, via controle *Bluetooth*, modificar a posição do *stickman* na tela, podendo também expandi-lo ou diminuí-lo de forma a realizar um ajuste da posição inicial do movimento e da distância relativa de execução.

D. Comunicação Bluetooth

A navegação do usuário na interface se dará por meio do controle *bluetooth*, será possível a escolha da atividade a ser executada e a entrada no menu de calibração, onde poderá ser feito o ajuste manual do *stickman*.

O controle contará com 7 botões e será implementado com um microcontrolador *MSP430* e um módulo de transmissão *bluetooth*, que comunicará com a *Raspberry* de forma serial. O algoritmo de recepção é capaz de parear com o microcontrolador e recebe os comandos de acordo com o botão apertado, tais comandos serão utilizados futuramente em um processo que se comunicará via *signals* com os demais processos, o que possibilitará a interação do usuário com o sistema. Tais funcionalidades são implementadas com a biblioteca *BlueZ*.

V. CONCLUSÃO

Um equipamento como o descrito neste trabalho possibilitaria à fisioterapeutas um acompanhamento mais efetivo ao tratamento de seus pacientes, por possuir a sua disposição uma série de dados sobre como os movimentos foram executados. O paciente terá maior segurança ao fazer os exercícios em casa sem o acompanhamento e eventualmente conseguirá corrigir seus movimentos. Desta forma, os tratamentos podem apresentar os resultados esperados mais rapidamente. Além disso, o baixo custo torna o produto acessível e atrativo frente às outras alternativas ao problema.

REFERÊNCIAS

- [1] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh, “Realtime multi-person 2d pose estimation using part affinity fields,” in *CVPR*, 2017.

- [2] Caroline Chan, Shiry Ginosar, Tinghui Zhou, and Alexei A Efros, "Everybody dance now," *arXiv preprint arXiv:1808.07371*, 2018.
- [3] Hanbyul Joo, Tomas Simon, Xulong Li, Hao Liu, Lei Tan, Lin Gui, Sean Banerjee, Timothy Scott Godisart, Bart Nabbe, Iain Matthews, Takeo Kanade, Shohei Nobuhara, and Yaser Sheikh, "Panoptic studio: A massively multiview system for social interaction capture," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.
- [4] Antonio Bo, Mitsuhiro Hayashibe, and Philippe Poignet, "Joint angle estimation in rehabilitation with inertial sensors and its integration with kinect," in *EMBC: Engineering in Medicine and Biology Conference*, 2011, pp. 3479–3483.
- [5] Stepan Obdrzalek, Gregorij Kurillo, Ferda Ofli, Ruzena Bajcsy, Edmund Seto, Holly Jimison, Michael Pavel, et al., "Accuracy and robustness of kinect pose estimation in the context of coaching of elderly population," in *Engineering in medicine and biology society (EMBC), 2012 annual international conference of the IEEE*. IEEE, 2012, vol. 28, pp. 1188–1193.