

① O primeiro passo a ser tomado é a obtenção da matriz de homografia. Para isso, pode-se utilizar a limitização automática do método de Otsu. Como a maior parcela da imagem corresponde ao campo, quase uniformemente verde, essa limitização tornaria o campo numa região branca (assume-se que a imagem já tenha sido convertida para grayscale antes desse processo). Essa região será o maior contorno da imagem, assim, pode-se encontrar as coordenadas que definem o campo ao se encontrar as coordenadas do maior contorno. Assim, consegue-se obter a matriz de homografia pela função `findHomography` da OpenCV. Esse procedimento assume que haverá um momento sem objetos e pessoas em campo e que o campo inteiro, pelo menos grande parte, estará visível. Caso a câmera não se mova ou dê zoom, não há necessidade de recalcular a homografia.

Falta então identificar as coordenadas da área do gol. Para isso, seria possível utilizar métodos de detecção de bordas e quinas, como o Harris detector ou Hough transform. Com base nessas características, encontrar os contornos por elas formados e obter os retângulos, ou paralelogramos, em campo. Fazendo uma limitização de área de contorno e posição relativa na imagem (caso a câmera seja fixa), pode-se obter as coordenadas das quinas da área do gol.

Com base nas posições de alvo encontradas e na matriz de homografia, que também pode ser calculada com a área do gol obtida, pode-se realizar a projeção perspectiva para alinhar os símbolos da forma mais "natural" para com a pose da câmera.

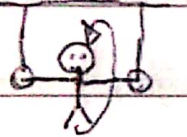
○ símbolo deveria ser rotacionado  $180^\circ$  antes de ser projetado na área de gol mais longe da câmera, que corresponde à área com menor coordenada vertical, devido à escala utilizada para imagens.

○ método proposto se baseia então em encontrar as regiões de projeção e matriz de homografia antes da transmissão da partida e assim assume que as câmeras se mantêm fixas e sem zoom. Dessa forma, evita-se tratamento de oclusão e somente é necessário calcular posição alvo e homografia uma única vez. Tomando o método rápido e robusto para essas condições especificadas, uma vez que sempre haverá formas retangulares e o campo será o melhor "objeto".

(2) Descritores como o SIFT dependem de informações locais, sendo baseados em computador histogramas (gradientes) através de janelas deslizantes. As logos apresentadas são bastante minimalistas, apresentando composição majoritariamente por cores sólidas e baixa variação de estrutura. Dessa forma as logos apresentam pouca informação, baixa entropia, e, uma vez que são imagens sintéticas, não apresentam ruído, logo o cálculo de gradientes é prejudicado. Uma vez que os gradientes são prejudicados, haverá poucos descritores realmente representativos, o que torna o casamento de imagens difícil.



⑤ O problema é análogo ao fenômeno apresentado quanto ao aperture problem exemplificado pelo "barber pole". Tomando-se as manobras de ginástica em barras e argolas, pode-se perceber que o movimento de rotação 3D será mal interpretado pelo algoritmo. Ao estimar o fluxo ótico por variação de intensidade, os vetores obtidos podem não representar o movimento real do atleta. Como por exemplo no desenho abaixo. O atleta gira em torno do eixo formado por seus braços,



⇒



Porém devido à posição da câmera o algoritmo não considera a rotação dos braços, já que não variam de intensidade, e provavelmente a captação de movimento perto do tronco seria difícil, supondo que o uniforme seja uma cor sólida. Assim somente seria estimado a cabeça indo para baixo e os pés para cima, porém não representando a rotação verdadeira.

⑦ 
$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$
 é uma transformação afim

$$\begin{bmatrix} wx' \\ wy' \\ w \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

$$b = \underbrace{A}_{\text{descontido}} x$$

pode ser escrito como (pontos)  $\Rightarrow$

$$\begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x'_1 x_1 & -x'_1 y_1 & -x'_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y'_1 x_1 & -y'_1 y_1 & -y'_1 \\ \vdots & & & & & & & & \\ x_n & y_n & 1 & 0 & 0 & 0 & -x'_n x_n & -x'_n y_n & -x'_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -y'_n x_n & -y'_n y_n & -y'_n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ c \\ c \\ c \\ e \\ c \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$2n \times 9$   
 $A$

$9 \times 1$      $2n \times 1$

Obtém-se  $h$  por  $\min_h \|Ah - 0\|^2$ ;  $\|h\|^2 = 1$

isso pode ser resolvido com decomposição em valores singulares: o autovetor correspondente ao menor valor singular (idealmente nulo) é o vetor  $h$

O sistema montado apresenta 9 equações, porém o vetor  $h$  apresenta 2 elementos nulos e um valor já determinado: 1. Dessa forma o sistema tem 3 graus de liberdade a menos. Considerando a restrição de norma, "perde-se" outro grau de liberdade, assim, necessita-se de no mínimo 5 pontos para resolver o sistema e obter a transformação afim



⑧ Pode-se pensar nos nódulos como blobs, dessa forma, pensa-se imediatamente em filtros LOG ou DoG.

Dessa forma, constrói-se o espaço de escala por meio da pirâmide gaussiana, utilizando DoG, ou seja subtração de gaussianas por  $DoG = G(x, y, \sigma_1) - G(x, y, \sigma_2)$ , buscando utilizar a melhor aproximação, ou seja,  $\sigma_1 = \sigma/\sqrt{2}$ ,  $\sigma_2 = \sigma\sqrt{2}$ . Ainda, buscar-se-ia 99% de suporte, o que ocorre quando a máscara utilizada tem tamanho 3 $\sigma$ .

Ao final, tendo-se variado  $\sigma$ , ter-se-á imagens somente com informações relevantes, blobs. Basta então contar o número de blobs (nódulos) baseado nos contornos encontrados, tendo-se em mente um threshold de área de contorno, de forma a evitar que o cérebro/crânio seja considerado como um grande blob (nódulo).

b) Supondo-se ter amostras suficientes de tomografias, com e sem presença de nódulos, todas já tendo sido rotuladas perfeitamente, pode-se então treinar uma rede convolucional (CNN) de forma supervisionada.

Para o treinamento, é ideal utilizar backpropagation e assim, deve-se inicializar os pesos da rede aleatoriamente, idealmente utilizando a inicialização de Xavier and He Normal (He et al). A determinação dos hiperparâmetros da rede, como número de camadas, tamanho dos filtros e funções de ativação, devem ser feito por meio de validação cruzada, utilizando processo de K-folds. O conjunto de dados obviamente deve ser subdividido em conjunto de treino e teste, com proporção maior de dados de treino.

A rede deveria ter a seguinte lógica geral: uma camada, ou mais, convolucional para detecção de blobs; uma rede convolucional posterior para computar o número de blobs; e por fim uma NN com ativação linear, ou ReLU, para retornar o número de nódulos. Pode-se pensar então que essa rede é formada por duas subredes: uma identifica a presença de nódulos e a outra conta o número de nódulos reportados pela rede anterior.

c) O uso de diversas camadas implica em filtrar parâmetros cada vez de mais alto nível. Ou seja, espera-se que a k-ésima camada identifique

peculiaridades como cor, posição, textura, etc. O problema em questão não deve ser específico, deseja-se contar o número de nódulos e não o número de nódulos em partes arbitrárias do cérebro ou nódulos de aspecto arbitrário. Dessa forma, o uso de muitas camadas pode ocasionar em um overfitting de forma que a rede não conta todo e qualquer nódulo, mas apenas aqueles que seguem um padrão por ela aprendido. Em contexto de saúde, isso é um grande problema. Além disso, o esforço computacional para treinar redes com muitas camadas é muito grande e, provavelmente, desnecessário para o problema.