

Computer Vision Project Report

Algorithm

To detect the 6 things I have used multiple methods.

1. To detect the id of the **pokemon** is a classification task and I used KNN with $k = 3$ to do it. The feature of choice was HoG and a suitable 'cellsize' was chosen after some trial and error. Once all the training features were found out from a cropped input images, I trained the model and stored it. Afterwards for every test image I create similar vector of HoG features and used the created feature vector to predict the label of the pokemon. HoG was used as it is best way to maintain and exploit the orientation information of a pokemon.
2. To detect the **center** I first found out the left lower point of the semicircle. This was done by first cropping the image to a suitable size around the left lower part of the semi-circle and then detecting a point of interest (using the most important point using surf feature). This was done as we have highest probability of getting the required point as the top most surf point. This is because it is surrounded by points of similar color and hence becomes a very important point. Then using the geometrical knowledge of pixels I found the other end (mirror image if axis is drawn in the center). Then the average of these points was taken to calculate the center of the circle. The factor with which the image was cropped is then added to this co-ordinates and the final actual co-ordinates are returned.
3. To detect the **level** I first crop the image to a suitable size and then try and detect a circle form the image. The image is converted so that the circle is bright as compared to the remaining image and then I detect bright circles in the image. The top most important circle is chosen as the most important one and its center is calculated. The factor with which the image was cropped is then added to this co-ordinates and the final actual co-ordinates are returned.
4. Similar logic is used for the remaining 3 entities – **CP, HP, Stardust**. Since it is text detection I first crop the suitable portion from the image and then detect if there is an any text content in that part of the image. MSER functions are used for this purpose. Once the text area is found, individual bounding boxes are generated for each character and these are then further cropped. The label is then attached to these individual images and this process is repeated for all the images in the training data. This is the way we create a training set. Now again, The feature of choice here was HoG and a suitable 'cellsize' was chosen after some trial and error. Using KNN for $K = 3$, I created a classification model for the digits. For test data similar process is done and HoG features are found out from the test images to create a test dataset. The label for each image is then predicted and all the labels are combined to get the actual value of the feature. The bounding boxes are sorted along the x axis to maintain correct ordering of all the images and all characters.

Experiments

I tried out a lot many combinations of algorithms and features for this project. It was a good learning experience.