The effectiveness of eye contour features in gaze estimation

Xiaodong Ning

September, 2023

1 Abstract

We proposed a naive gaze estimation method and showed the effectiveness of the eye shape feature in this article. It first extracted feature vectors based on the eye shape segmentation results and the histogram of oriented gradients (HOG). And then a mutiple-output Support Vector Regression (SVR) model was trained with the features. The experiment in the part of Multi-view Gaze Dataset (CVPR' 14) showed that the combination of the 2 types of features outperformed each of them alone. It was indicated that the segmentation result could be used to improve the feature descriptor based prediction. Therefore, a further exploration of the segmentation could be made (e.g., Deep Learning based segmentation) for the purpose of more accurate gaze estimation.

2 Method

2.1 Eye contour feature

The segmentation result of the eye shape is used as a prediction feature. The method generates a mask (a binary image) of the eye, resizes the mask and flattens it into the feature vector. We use Chan-Vese segmentation [1] for the segmentation task. [1] assumes that the average intensities of the object region and non-object region should be different. So tt is based on the level sets that are evolved iteratively to minimize an energy, which is defined by:

$$E = \int_{inside(C)} |u_0 - c_1|^2 dxdy + \int_{outside(C)} |u_0 - c_2|^2 dxdy$$

where C is the object contour, u_0 is the image intensity, c_1 and c_2 are the average pixel intensity values inside/outside the contour respectively. More regularizing terms are added, like the length of C and the area inside C, for a better performance:

$$E = \mu length(C) + \upsilon area(inside(C)) + \lambda_1 \int_{inside(C)} |u_0 - c_1|^2 dxdy + \lambda_2 \int_{outside(C)} |u_0 - c_2|^2 dxdy$$

where μ, υ, λ_1 and λ_2 are fixed parameters.

Thus, the segmentation is converted to a minimizing problem. A binary function, Heaviside function $H(x) = sgn(x)$, is used to indicate the length and inside area of the object.And the energy could be expressed in a more effective way:

$$E = \mu \int_\Omega |\nabla H(\phi)| dxdy + \upsilon \int_\Omega H(\phi) dxdy + \lambda_1 \int_\Omega |u_0 - c_1|^2 H(\phi) dxdy + \lambda_2 \int_\Omega |u_0 - c_2|^2 (1 - H(\phi)) dxdy$$

where Ω is the domain of the whole image, and ϕ is a function on Ω, $\phi(x, y)$, and the object contour can be the set of points, $(x, y) \in \Omega : \phi(x, y) = 0$.

Therefore the gradient descent of the energy E should be:

It iterates the contour ϕ from the initialization status to a locally optimal solution.

2.2 Image depth feature

References

[1]

[2] TODO.