

Scanned images of historical documents often suffer from bleed-through, which refers to the ink on one side seeping through the paper and appearing on the other side. In this paper, a new conditional random field (CRF)-based method is proposed to remove the bleed-through from the scanned images of historical images. The proposed method only requires the scanned image of one side, referred as a blind method. In general, the scanned historical document image is composed of three components: foreground, bleed-through, and background. By assuming Gaussian distributions of the three components, the proposed method establishes conditional probability distribution (CPD) models of the three components first. The parameters of the component CPD models are estimated based on an initial segmentation of the input image. Then, CRFs are used to capture the relations between observed pixels in the scanned image and the corresponding labels as well as the spatial relation between the adjacent labels. The belief propagation algorithm is used to calculate the probabilities of different labels for each pixel. Once the labeling is completed by choosing the most possible label for each pixel, the bleed-through component is removed from the input historical image by a random-filling inpainting algorithm. Experimental results on the real data set show that the proposed method preserves the foreground component very well and removes the bleed-through effectively.