Recognition and understanding of text in scene images is an important and challenging task. The importance can be seen in the context of tasks such as assisted navigation for the blind, providing directions to driverless cars, e.g. Google car, etc. Other applications include automated document archival services, mining text from images, and so on. The challenge comes from a variety of factors, like variable typefaces, uncontrolled imaging conditions, and various sources of noise corrupting the captured images. In this work, we study and address the fundamental problem of recognition of characters extracted from natural scene images, and contribute three holistic strategies to deal with this challenging task.

Scene text recognition (STR) has been a known problem in computer vision and pattern recognition community for over two decades, and is still an active area of research owing to the fact that the recognition performance has still got a lot of room for improvement. Recognition of characters lies at the heart of STR and is a crucial component for a reliable STR system. Most of the current methods heavily rely on discriminative power of local features, such as histograms of oriented gradient (HoG), scale invariant feature transform (SIFT), shape contexts (SC), geometric blur (GB), etc. One of the problems with such methods is that the local features are rasterized in an ad hoc manner to get a single vector for subsequent use in recognition. This rearrangement of features clearly perturbs the spatial correlations that may carry crucial information vis-à-vis recognition. Moreover, such approaches, in general, do not take into account the rotational invariance property that often leads to failed recognition in cases where characters in scene images do not occur in upright position. To eliminate this local feature dependency and the associated problems, we propose the following three holistic solutions:

The first one is based on modelling character images of a class as a 3-mode tensor and then factoring it into a set of rank-1 matrices and the associated mixing coefficients. Each set of rank-1 matrices spans the solution subspace of a specific image class and enables us to capture the required holistic signature for each character class along with the mixing coefficients associated with each character image. During recognition, we project each test image onto the candidate subspaces to derive its mixing coefficients, which are eventually used for final classification.

The second approach we study in this work lets us form a novel holistic feature for character recognition based on active contour model, also known as snakes. Our feature vector is based on two variables, direction and distance, cumulatively traversed by each point as the initial circular contour evolves under the force field induced by the character image. The initial contour design in conjunction with cross-correlation based similarity metric enables us to account for rotational variance in the character image.

Our third approach is based on modelling a 3-mode tensor via rotation of a single image. This is different from our tensor based approach described above in that we form the tensor using a single image instead of collecting a specific number of samples of a particular class. In this case, to generate a 3D image cube, we rotate an image through a predefined range of angles. This enables us to explicitly capture rotational variance and leads to better performance than various local approaches.

Finally, as an application, we use our holistic model to recognize word images extracted from natural scenes. Here we first use our novel word segmentation method based on image seam analysis to split a scene word into individual character images. We then apply our holistic model to recognize individual letters and use a spell-checker module to get the final word prediction.

Throughout our work, we employ popular scene text datasets, like Chars74K-Font, Chars74K-Image, SVT, and ICDAR03, which include synthetic and natural image sets, to test the performance of our strategies. We compare results of our recognition models with several baseline methods and show comparable or better performance than several local feature-based methods justifying thus the importance of holistic strategies.
Educational Career:
Bachelor's of Engineering, BS, 2000, University of Engineering and Technology, Lahore, Pakistan
Master's of Computer Science, MS, 2004, University of the Punjab, Lahore, Pakistan
Master's of Computer Science, MS, 2014, University of Central Florida

Committee in Charge:
Hassan Foroosh, Chair, Computer Science
Charles E. Hughes, Computer Science
Gita R. Sukthankar, Computer Science
Rudolph P. Wiegand, Institute for Simulation and Training
Hae-Bum Yun, Civil, Environmental, and Construction Engineering

Approved for distribution by Hassan Foroosh, Committee Chair, on June 18, 2016.

The public is welcome to attend.