Unrolling the Shutter: CNN to Correct Motion Distortions

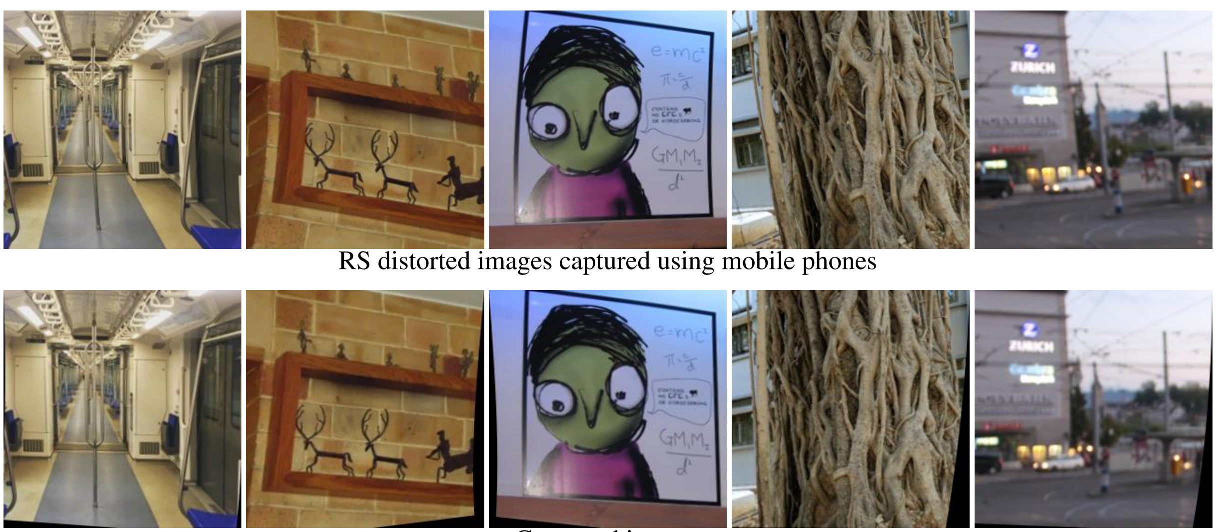
Supplementary PDF

Vijay Rengarajan[1*], Yogesh Balaji[2†], A.N. Rajagopalan[3]
[1,3]Indian Institute of Technology Madras, [2]University of Maryland
*vijay.ap@ee.iitm.ac.in

This supplementary document is arranged as follows: (i) additional results for real images and from the test sets, (ii) some failure cases, (iii) description of human perception rating survey, and (iv) analysis of RowColCNN banks.

1. Additional Results

We show additional results of our correction method using RowColCNN for the images that we captured using our mobile phones in Fig. S1. Image (a) is a common urban scene image of the inside of a train with many lines in the scene and our method performs well in correcting the RS distortion. Images (b) and (c) are more challenging in that there are structures other than straight lines; even in these cases, our method works well. Image (d) is not a regular urban image but an image of a banyan tree having many natural curved structures in its trunk. It has a skew RS effect since it is taken from a moving vehicle. Our method corrects the skew in this case too despite being not specifically trained in the presence of natural curvatures. The last image (e) contains some blur in addition to the RS distortion, and our method corrects the slanting effect well.



RS distorted images captured using mobile phones

Corrected images

(a) (b) (c) (d) (e)

Figure S1. Results of RowColCNN for real images.

We now show additional results from the test sets. Fig. S2 shows RS distorted images and their corresponding output images from the urban test set. The CNN is able to learn curvature correction even for the case of rotated images (slanted away from the vertical) as shown in (b) and (c). Fig. S3 shows corrections from the face test set. RowColCNN is able to correct diverse types of faces, and it is also robust to face markings as shown in the third column as well as occluding objects as shown in (d).

†This work was done when the second author was studying at IIT Madras.

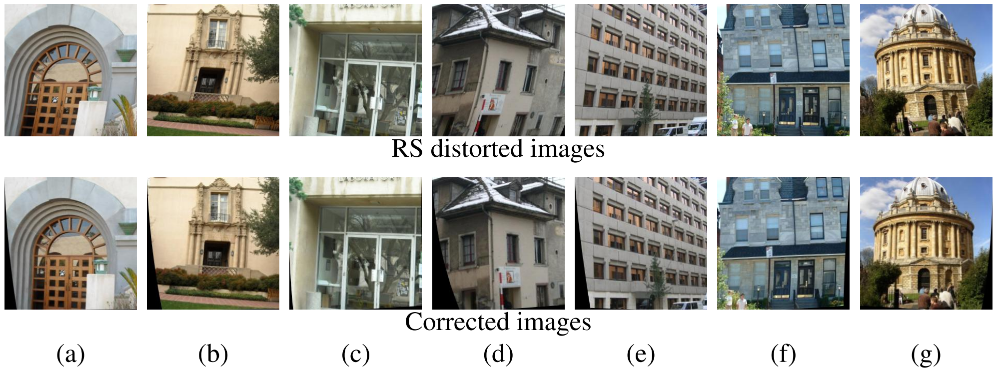


Figure S2. Results of RowColCNN from urban testing set.

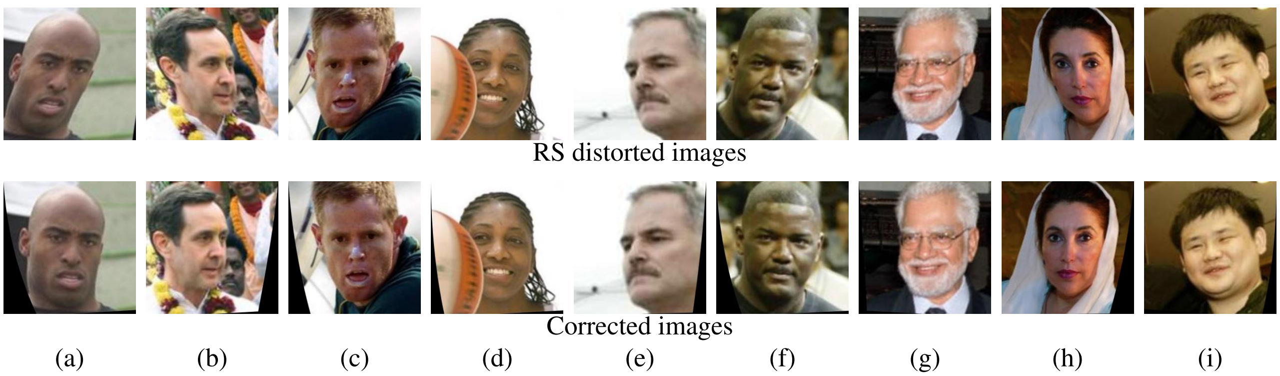


Figure S3. Results of RowColCNN from face testing set.

Figure S4. Failure Results of RowColCNN from urban testing set.

2. Failure Results

In Fig. S4, we show some results of RowColCNN in which the RS distortions are not corrected properly. In (a), a number of branches and shadows in the scene causes RowColCNN to not correct the RS effect fully, leaving a small residual curvature in the output image. In (b) and (c), small residual distortion remains in the corrected image when compared with the ground-truth image, and yet, the outputs are not visually displeasing to the viewer.

3. Human Perception Survey

Even though the error in motion estimation and the comparison of PSNR help to quantify and compare different RS correction algorithms, we additionally sought a human rating for the corrected images. There might be cases where the estimated motion could be different from that of the ground-truth motion with the corrected image being still visually plausible as shown in the last two examples of the previous section. To address this behavior, we surveyed 63 humans (42 for face class and 21 for building class) for their judgment based on what they visually perceive in the final output.

The survey consisted of 30 questions each in urban scene class and face class. Each question had four images that had to be rated, namely, (i) the ground-truth GS image, (ii) the distorted RS image, (iii) RS corrected image using RowColCNN (our method), and (iv) RS corrected image using the competing methods [13] for urban class and [4] for face class. Every image could be assigned one of the four values, 1 to 4, with 1 denoting the better among the given four.

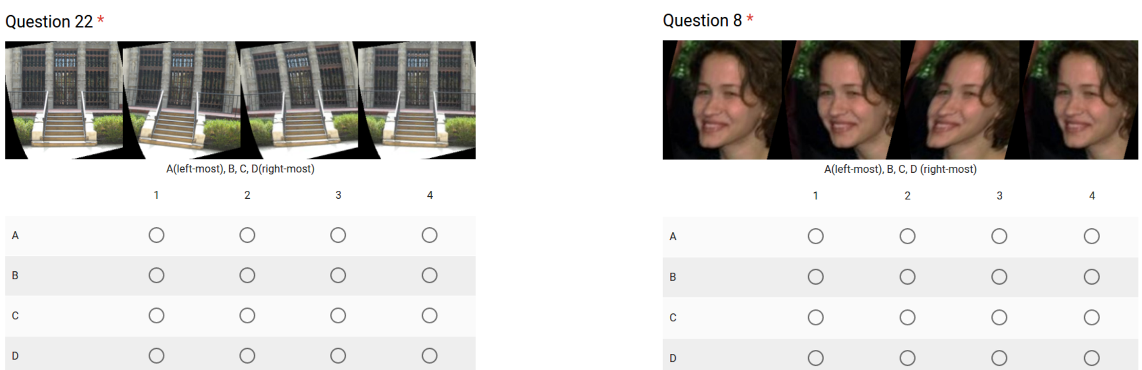
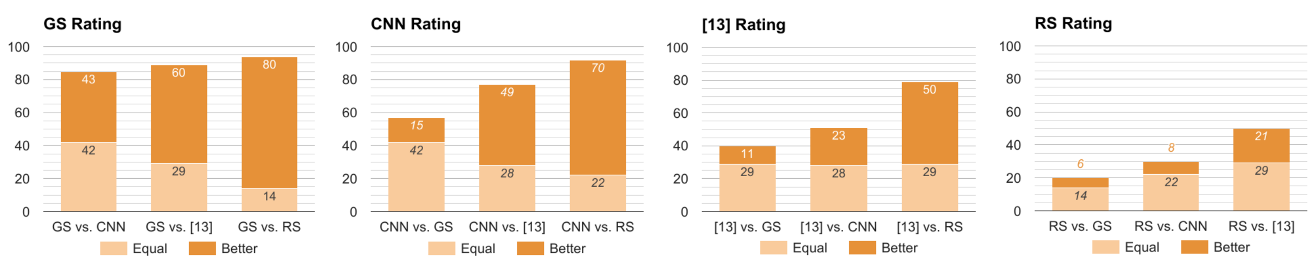
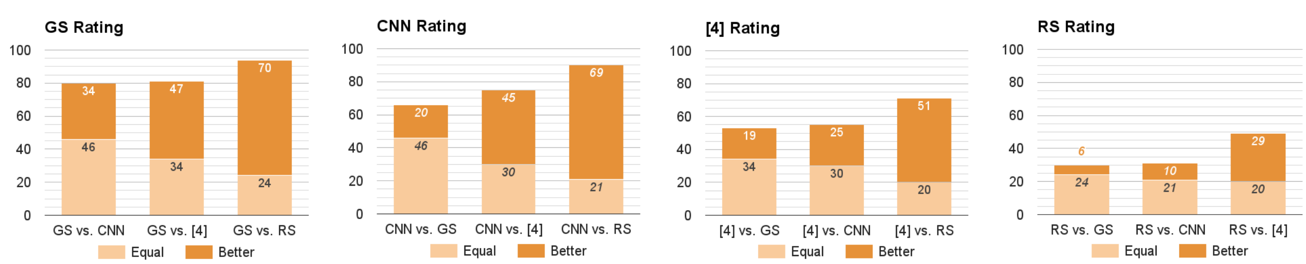


Figure S5. *Human Perception Survey*: Sample questions provided to the participants.



(a) Rating for urban images of different methods against the rest of the methods.



(b) Rating for face images of different methods against the rest of the methods.

Figure S6. Human perception rating survey results.

Fig. S5 shows two sample questions from the survey. The order of the four images are randomized when presented to the users, and they are *not* given the label of images (i.e. GS, RS, or correction methods). Since the RS correction results in black borders due to warping, we combine those black regions from the corrected images and apply the same to all images. The participants have the choice to select one option (1, 2, 3, or 4) for each of the images A, B, C, and D. The same rating can be chosen for more than one image if the user feels that they look visually similar. For example, the user can rate two images with the same value of 1; in this case, the next choice should continue with 2 and not 3, i.e. the ranks have to be consecutive. The images used for the survey, their random ordering, and user responses are provided in the supplementary zip file.

The survey results are given in Fig. S6. Each plot shows how *equal* or *better* a particular image out of the four labels performs against the other three images. The percentages of both *equal* rating and *better* rating are shown in the plots. For instance, in the *CNN Rating* plot of Fig. S6(a), one can read that the output image of the CNN is rated equally 28% of the time against that of [13], and it is rated above [13] 49% of the time by the users. Our inferences from the survey results are as follows:

- The undistorted GS image is preferred equal to or above all others at least 80% of the time against all other three images in both the urban and face classes. It is preferred 94% of the time against the RS image and it is expected.
- Our CNN is better than the competing methods [13] and [4] in the following aspects:
  (a) It performs equally or better than the original GS image 57% and 66% of the time as against 40% and 53% of the time for [13] and [4] for urban and face datasets, respectively.
  (b) It is preferred more than [13] and [4] (ignoring *equal* performance) 49% and 45% of the time, respectively. If the *equal* rating is included, its preference jumps to 77% and 75%, respectively.
- The corrected images from our method as well as competing methods are rated better than the RS image which is to be expected.

4. RowColCNN Analysis

In this section, we analyze the RowColCNN based on its bank responses and translation invariance property.

4.1. Responses of Row and Column Banks

We wish to observe the excitation of row and column kernel banks of RowColCNN for images containing differently oriented edges. To achieve this, we look at the response vector at the end of the two banks for different input images. We calculate the number of values above a particular threshold value to get a sense of how large is the contribution of each bank. We obtain a plot by varying this threshold which tells how fast or slow the magnitude of values decreases. We created images of black and white strips at different angles from vertical to horizontal for this purpose. Fig. S7 shows the illustration of the formula corresponding to the plot.

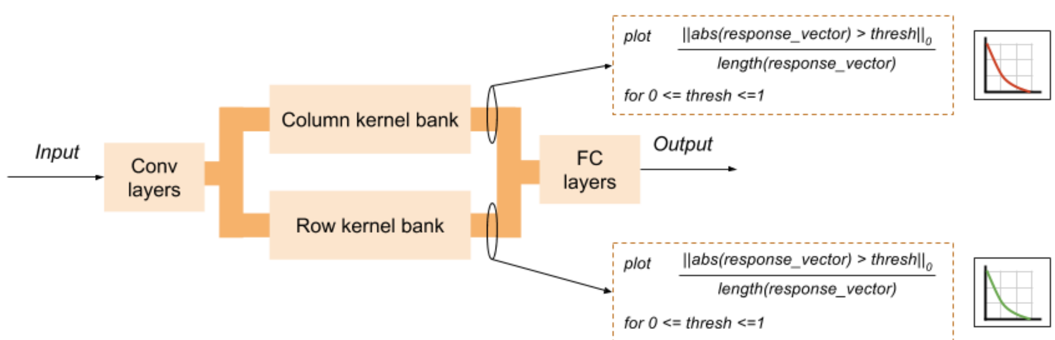


Figure S7. *Responses of Row and Column Banks*: We tap the response vector of each bank, and plot the ratio of number of elements greater than a threshold for a range of threshold values.

Fig. S8 shows the response plots corresponding to row and column kernel banks. We make the following observations:

- Both column and row banks are excited to see vertical lines more than horizontal lines. This can be seen from the dropping rate of the plots of both banks where the darker plots corresponding to more horizontally-oriented lines fall faster than the lighter plots (red or green).

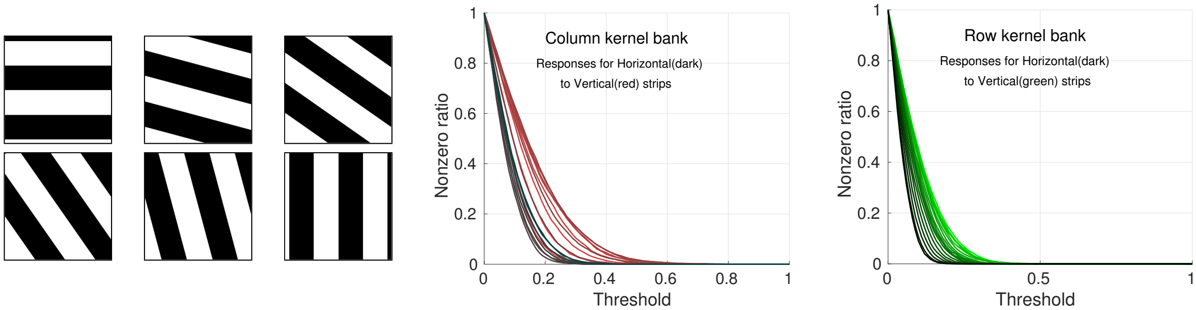


Figure S8. *Responses of Row and Column Banks*: (Left) Sample input images of strips at different angles. (Right) Response at the bank output.

- The excitation level is more for the column bank than for the row bank for a line at any particular angle. The plots for the row bank fall to zero much earlier than the corresponding plots for the column bank.
- The ratio of excitation between vertical and horizontal lines is higher for the column bank than for the row bank. This can be inferred from the width of the two plots.

4.2. Invariance under Global Translations

A *checksum* condition for any single-image RS motion estimation method is its consistency in estimating the same camera motion when different globally translated versions of the same image are provided as inputs, i.e. it should output the same motion for all globally translated versions of the same image. To verify this behavior, we generated 21 different globally translated versions of ten different zero-motion images, some of which are shown in Fig. S9 (left). We then allowed Row-ColCNN to predict the motion trajectory for each of these images. The expectation is to have zero predicted motion for all images.

Fig. S9 (right) shows the plots of the mean motion trajectories and their standard deviations of all the predicted motions. It can be seen that the estimated trajectory is very close to zero for both translations and rotations. (Note that the axis range in the plot is [-1,1] pixels for translation and [-0.5,0.5] degrees for rotation.) Hence, we can say that RowColCNN is invariant under global translations.

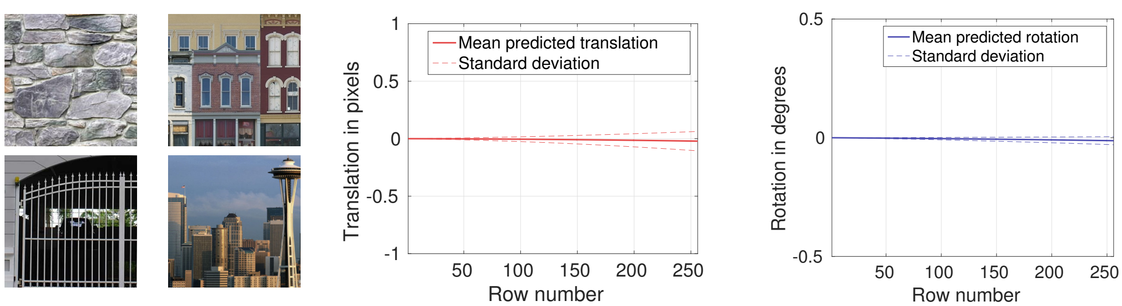


Figure S9. *Invariance under Global Translations*: (Left) Sample images used to check translation invariance. (Right) Mean and standard deviation of the estimated motion trajectories.