Radially-Distorted Conjugate Translations Supplementary Material

James Pritts[1] Zuzana Kukelova[1] Viktor Larsson[2] Ondřej Chum[1]
Visual Recognition Group, CTU in Prague[1] Centre for Mathematical Sciences, Lund University[2]

A. Transfer Error

The scene plane is tessellated by a 10x10 square grid of points, denoted $\{\mathbf{X}_i\}$, with a 1 meter spacing between adjacent points. Suppose that $\mathbf{y} \leftrightarrow \mathbf{y}'$ is an undistorted point correspondence induced by the conjugate translation $\mathtt{H}_{\mathbf{u}} = [\mathtt{I}_3 + \mathbf{u}\mathbf{l}^\top]$ in the imaged scene plane (here we assume that $s_i^{\mathbf{u}} = 1$ since we speak about an individual point correspondence).

Points $\{\mathbf{X}_i\}$ are translated by 1 meter on the scene plane in the direction of translation induced by the preimage of the point correspondence $\mathbf{y} \leftrightarrow \mathbf{y}'$ giving the translated grid $\{\mathbf{X}_i'\}$. The purpose of constructing the grid and it's translation is to uniformly cover the scene plane that the camera images in its field of view. In this way, the accuracy of the conjugate translation and lens-distortion parameter estimation can be measured across most of the image. The conjugate translation $\mathtt{H}_{\mathbf{u}}$ is not used directly because the magnitude of translation may span the extent of the scene plane, so applying it to the tessellation would transform the grid out of the field of view.

Let the camera be parameterized by the camera matrix $\mathtt{P} = (\mathtt{AH})^{-1}$ (see Sec. 5.4 for the definition of the camera matrix) that pointwise maps the scene plane Π to the imaged scene plane π and division model parameter λ. The preimages of the undistorted point correspondence $\mathbf{y} \leftrightarrow \mathbf{y}'$ in the scene-plane coordinate system is, respectively, $\beta\mathbf{Y} = \mathtt{P}^{-1}\mathbf{y}$ and $\beta'\mathbf{Y}' = \mathtt{P}^{-1}\mathbf{y}'$. The translation $\mathbf{t}$ of the preimages in the scene plane coordinate system is $\mathbf{t} = \mathbf{Y} - \mathbf{Y}' = (t_x, t_y, 0)^\top$.

Then $\|\mathbf{t}\|$ is the magnitude of translation between the repeated scene elements in the scene-plane coordinate system. Denote the homogeneous translation matrix $\mathtt{T}(\mathbf{t})$ to be the matrix constructed from $\mathbf{t}$ as

$$\mathtt{T}(\mathbf{t}) = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{12}$$

The translation of the grid points by unit distance is given by $\mathbf{X}_i' = \mathtt{T}(\mathbf{t}/\|\mathbf{t}\|)\mathbf{X}_i$. Recall from (1) that a conjugate translation has the form $\mathtt{PT}(\cdot)\mathtt{P}^{-1}$. Using (2), the conjugate translation of unit distance in the direction of point correspondences $\mathbf{y} \leftrightarrow \mathbf{y}'$ is

$$\begin{aligned} \mathtt{H}_{\mathbf{u}/\|\mathbf{t}\|} &= \mathtt{P}\mathtt{I}_3\mathtt{P}^{-1} + \mathtt{P}\begin{pmatrix} t_x/\|\mathbf{t}\| \\ t_y/\|\mathbf{t}\| \\ 1 \end{pmatrix}\begin{bmatrix} \mathtt{P}^{-\top}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \end{bmatrix}^\top \\ &= [\mathtt{I}_3 + \frac{\mathbf{u}}{\|\mathbf{t}\|}\mathbf{l}^\top]. \end{aligned} \tag{13}$$

The unit conjugate translation $\mathtt{H}_{\mathbf{u}/\|\mathbf{t}\|}$ can be written in terms of the conjugate translation $\mathtt{H}_{\mathbf{u}}$ induced by the undistorted point correspondence $\mathbf{y} \leftrightarrow \mathbf{y}'$ as

$$\begin{aligned} \mathtt{I}_3 + \frac{\mathbf{u}}{\|\mathbf{t}\|}\mathbf{l}^\top &= \mathtt{I}_3 + \frac{1}{\|\mathbf{t}\|}[\mathtt{I}_3 + \mathbf{u}\mathbf{l}^\top - \mathtt{I}_3] \\ &= \mathtt{I}_3 + \frac{1}{\|\mathbf{t}\|}[\mathtt{H}_{\mathbf{u}} - \mathtt{I}_3]. \end{aligned} \tag{14}$$

The derivation of (14) gives the form of transformation used in the transfer error $\Delta^{\text{xfer}}_{\text{RMS}}$ defined in Sec. 5.1, which maps from the undistorted points of the grid $\{\mathbf{x}_i\}$ to their translated correspondences $\{\mathbf{x}_i'\}$.

B. Computational Complexity.

Table B.1 lists the elimiation template sizes for the proposed solvers. The average time to compute the solutions for a given input for a solver is directly proportional to the elimination template size. The solvers are implemented in MATLAB and C++. Significantly faster implementations are possible with a pure C++ port. Still the solvers are sufficiently fast for use in RANSAC. The proposed solvers have an average solve time from 0.3 to 2 milliseconds.

$\mathtt{H}2.5\mathbf{lu}\lambda$	$\mathtt{H}3\mathbf{lu}s_{\mathbf{u}}\lambda$	$\mathtt{H}3.5\mathbf{luv}\lambda$	$\mathtt{H}4\mathbf{luv}s_{\mathbf{v}}\lambda$
14x18	24x26	54x60	76x80

Table B.1: Template sizes for the proposed solvers.

C. Extended Experiments

The extended real-data experiments in the following pages include (i) images with lesser radial distortion from consumer cameras and mobile phones; the images also demonstrate the proposed method's effectiveness on diverse scene content, (ii) images for very wide field-of-view lenses (8mm and 12mm), (iii) and an additional local-optimization experiment similar to the one in Fig. 6 on a GoPro Hero 4 image taken with its medium field-of-view setting, which further demonstrates the need for a minimal solver that jointly estimates lens distortion with affine-rectification to achieve an accurate undistortions and rectifications.

Figure C.1: *Narrow field-of-view and diverse scene-content experiments for* H2.5**lu**λ+LO. The proposed method works well if the input image has little or no radial lens distortion. This imagery is typical of consumer cameras and mobile phone cameras. The images are diverse and contain unconventional scene content. Input images are on the top row; undistorted images are on the middles row, and the rectified images are on the bottom row.

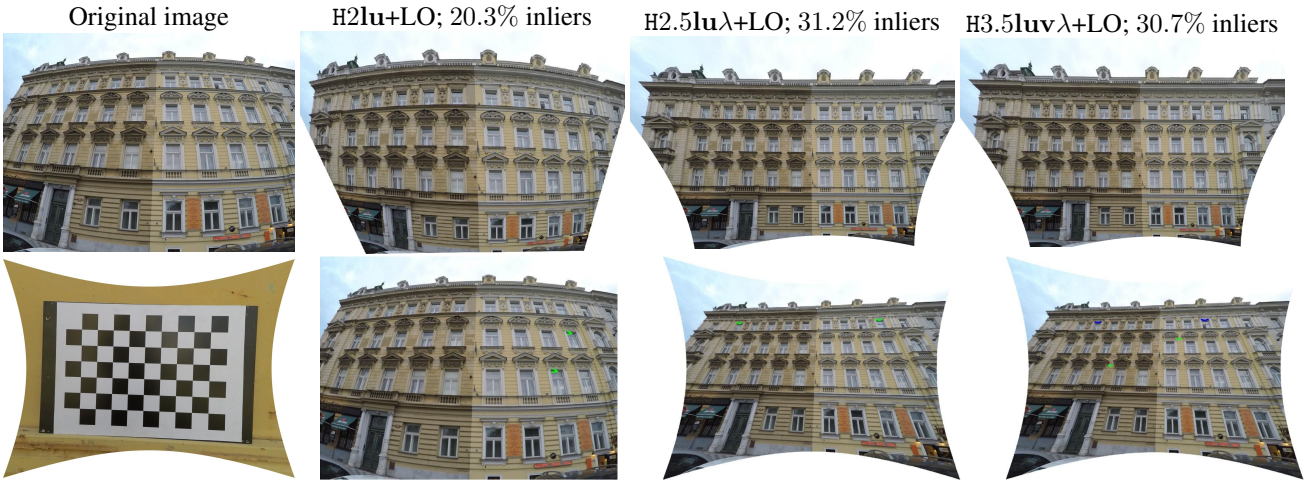

Figure C.2: *GoPro Hero 4 at the medium setting for different solvers.* Results from LO-RANSAC (see Sec. 4) for H2**lu**, which omits distortion, and the proposed solvers H2.5**lu**λ and H3.5**luv**λ. The top row has rectifications after local optimization (LO); The bottom row has undistortions estimated from the best *minimal* sample. LO-RANSAC fails from the poor initializations by H2**lu** (column 2). The proposed solvers in columns 3 and 4 give a correct rectification. The bottom left has a chessboard undistorted using the division parameter estimated from the building facade by H2.5**lu**λ+LO.

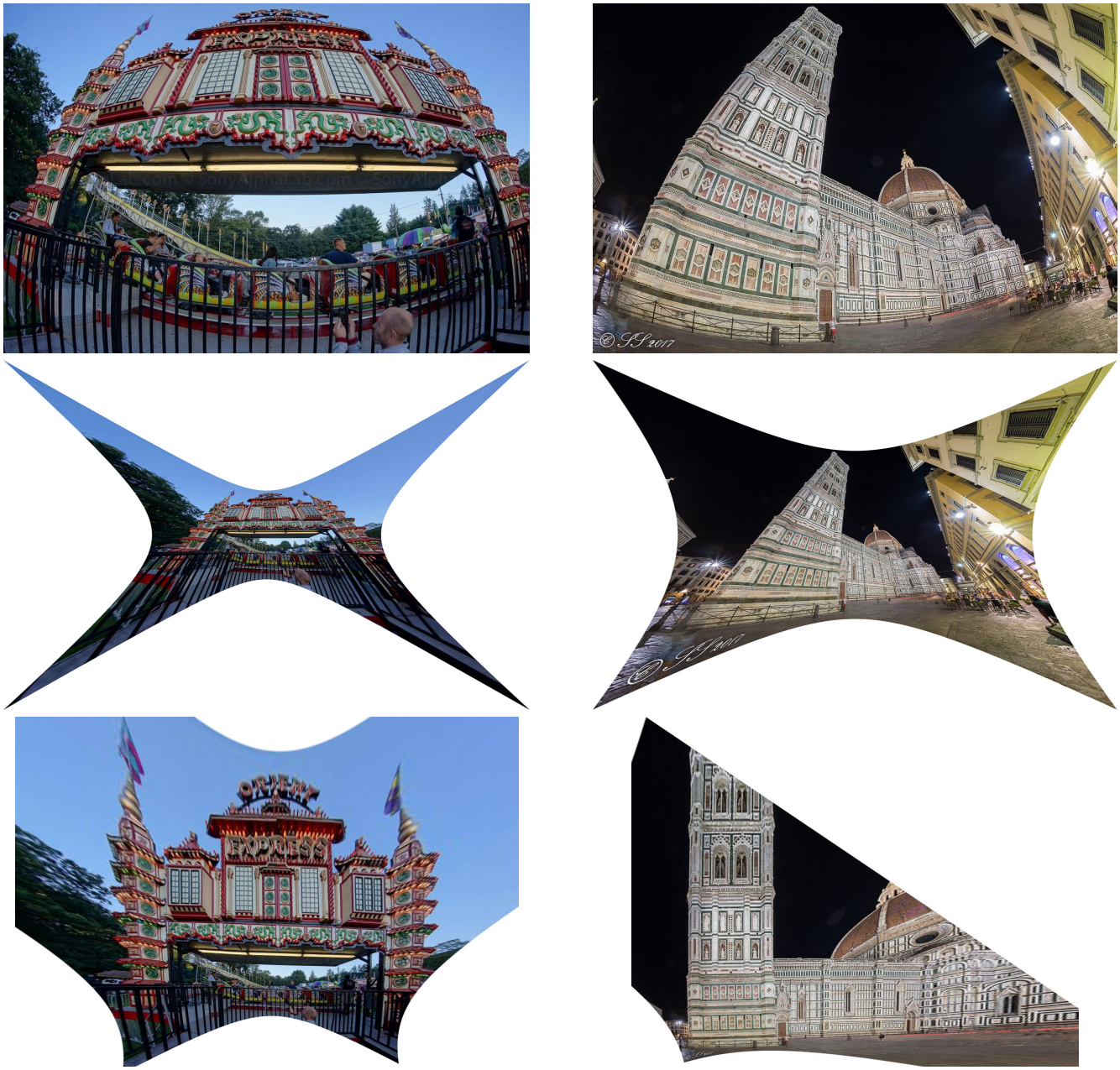

Figure C.3: *Very wide-angle images undistorted and rectified with* H2.5**lu**λ+LO. The left column is an image from an 8mm lens, and the right column is from a 12mm lens. The top row contains the input images; the middle row contains the undistorted images, and the bottom row contains the rectified images. The division model [7] used for radial lens distortion has only 1 parameter, which may impose limits for modeling extreme lens distortion.

Figure C.4: *Problem difficulty and method robustness* The input to the method is ungrouped affine-covriant features (top left). Common problems include missed detections of repeated texture, duplicate detections, and detections due to compression artifacts, all of which are visible in this example. The inliers with respect to the total number of detected features can be a very small proportion (top right). Still the method can estimate accurate undistortion (bottom left) and affine rectification (bottom right) even from a very sparse sampling of the inlying affine-covariant features on the scene plane (top right).