Joint Estimation of Deformable Motion and Photometric Parameters in Single View Video

Anna Hilsmann and Peter Eisert
Computer Vision and Graphics Group
Fraunhofer Heinrich Hertz Institute, Berlin, Germany

Overview
We present a method for joint deformation and illumination parameter estimation from monocular image sequences exploiting direct image information. We are particularly interested in augmented reality applications, where a new texture is rendered onto a moving and deforming surface in the original video in real-time. Realistic retexturing not only requires geometric registration but also photometric parameter retrieval for convincing illusion. Taking into account the photometric part by relaxing the brightness constancy assumption of the estimation not only allows realistic augmentation results but also improves spatial tracking.

Joint Motion and Illumination Estimation
We describe the spatial deformation of the surface in the image plane in a dense pixel displacement field and the photometric changes in a dense pixel intensity scale field and parameterize these fields with a deformable model that is presented as a planar triangulated regular 2D mesh with \( \nu \) vertices \( \mathbf{v}_i, (k = 1..\nu) \).

\[
\mathbf{D}(\mathbf{x}) = \sum_{i=1}^{\nu} \beta_i \mathbf{v}_i
\]

\[
S(\mathbf{x}) = \sum_{i=1}^{\nu} \beta_i \mu_i
\]

The resulting parameter vector modeling the deformable motion and illumination changes is then given by concatenating the parameters of each vertex:

\[
\Theta = \left( \delta \mathbf{v}_1, \ldots, \delta \mathbf{v}_\nu, \delta \mu_1, \ldots, \delta \mu_\nu \right)^T
\]

Generally, estimating the parameter vector leads to minimizing a cost function that consists of two terms:

\[
\hat{\Theta} = \arg\min_\Theta \left( E_a(\Theta) + \lambda \cdot E_S(\Theta) \right)
\]

\[E_D(\Theta) = \left\| \mathbf{J} \Theta - \mathbf{b} \right\|^2\]

The sparse structure of the Gauss-Newton approximation of the Hessian is exploited to solve the normal equations of the Linear Equation System. For a higher order interpolation, like B-splines or thin-plate splines can be formed in a similar way but the matrix is less sparse.

\[E_S(\Theta) = \left\| \mathbf{L} \Theta \right\|^2\]

\[\mathbf{L} = \text{Distance Weighted Laplacian Matrix}\]

Experimental Validation

RMSE Based Registration Evaluation
We calculate the Root Mean Squared Error (RMSE) between the synthetic image, generated from the computed geometric and photometric parameters, and the original current frame over all image pixels in the mesh region for several video sequences. Taking illumination parameters into account significantly reduces the mean RMSE over the entire sequence by up to 74%.

Point Based Registration Evaluation
We additionally calculated the distance between the estimated position of feature points to manually selected ground truth points. The mean difference between the estimated positions and the manually labeled ground truth position describes the geometric registration error. This additional evaluation approach is chosen to evaluate geometric registration accuracy separately from photometric registration. We can reduce the mean distance between the estimated and the ground truth position by approximately 40%.

Retexturing Results
The geometrical and photometric parameters are exploited to retexture a deforming surface in monocular images by establishing a shading map from the photometric parameters.