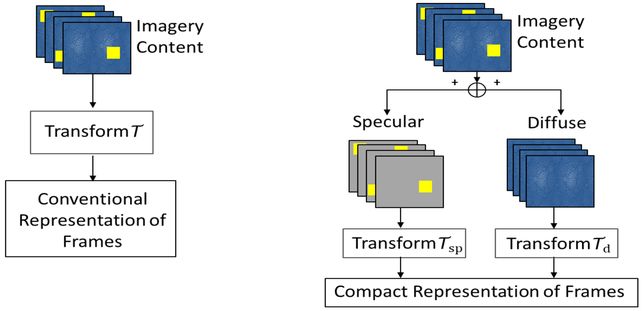
Rate-Distortion Driven Separation of Diffuse and Specular Components in Multiview Imagery

Maryam Haghighat, Reji Mathew, and David Taubman, *University of New South Wales, Sydney, Australia*

ICIP 2019
TAIPEI, TAIWAN

The Key Idea

We propose to decompose multi-view imagery into two additive parts which can be understood as diffuse and specular content. We choose distinct and different sparsifying transforms for the diffuse and specular contents and employ an R-D inspired measure as our optimization cost function to drive the decomposition based solely on compressibility.



Problem: separate $\mathrm{x}_{sp} = (x_{0,sp}, \ldots, x_{K,sp})^t$ from $\mathrm{x}_d = (x_{0,d}, \ldots, x_{K,d})^t$ such that total coding cost J^{tot} is minimized

$$\{\mathrm{x}_{sp}, \mathrm{x}_d\} = \underset{\{\mathrm{x}_{sp}, \mathrm{x}_d\}}{\operatorname{argmin}} J^{tot}(\mathcal{T}_{sp}\,\mathrm{x}_{sp}, \mathcal{T}_d\,\mathrm{x}_d)$$

subject to

$x_k = x_{k,d} + x_{k,sp}$, k=frame number

$\mathcal{T}_d$ is an inter plus intra view transform

$\mathcal{T}_{sp}$ is an intra view only transform

Coding Cost Function

$$J^{tot} = D^{tot} + \lambda L^{tot}$$

$$J^{tot} = \sum_{s,\mathbf{n}} J(y_{s,\mathbf{n}}) = \sum_{s,\mathbf{n}} D(y_{s,\mathbf{n}}) + \lambda L(y_{s,\mathbf{n}})$$

$y_{s,\mathbf{n}} \in \{\mathcal{T}_{sp}\,\mathrm{x}_{sp}, \mathcal{T}_d\,\mathrm{x}_d\}$

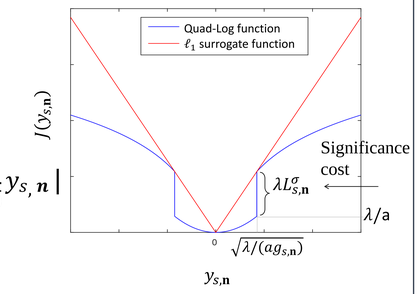
D= Quantization Distortion

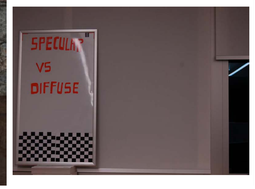
L= Coded Length

Problem Formulation

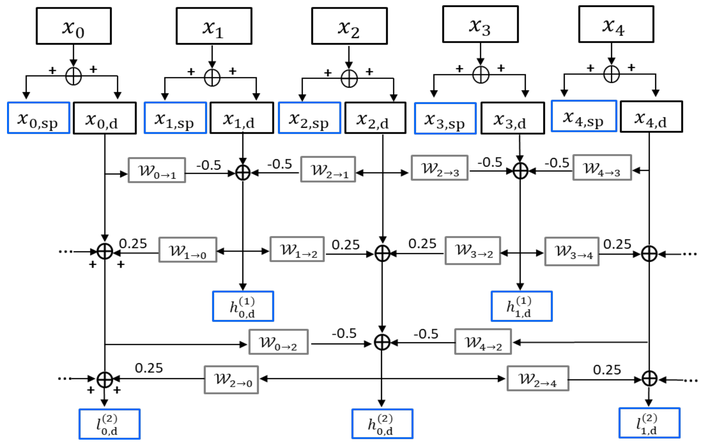
$$\{\mathrm{x}_{sp}, \mathrm{x}_d\} = \underset{\{\mathrm{x}_{sp}, \mathrm{x}_d\}}{\operatorname{argmin}} \sum_{s,\mathbf{n}} \left| \frac{\lambda(\frac{1}{a} + L^{\sigma}_{s,\mathbf{n}})}{\sqrt{\lambda/(a g_{s,\mathbf{n}})}} y_{s,\mathbf{n}} \right|$$

subject to $\mathrm{x} = \mathrm{x}_d + \mathrm{x}_{sp}$

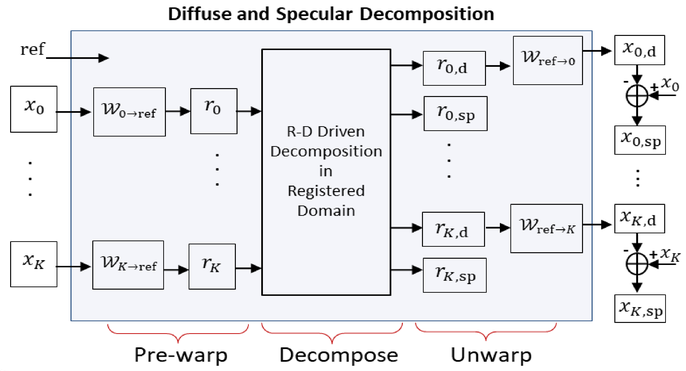




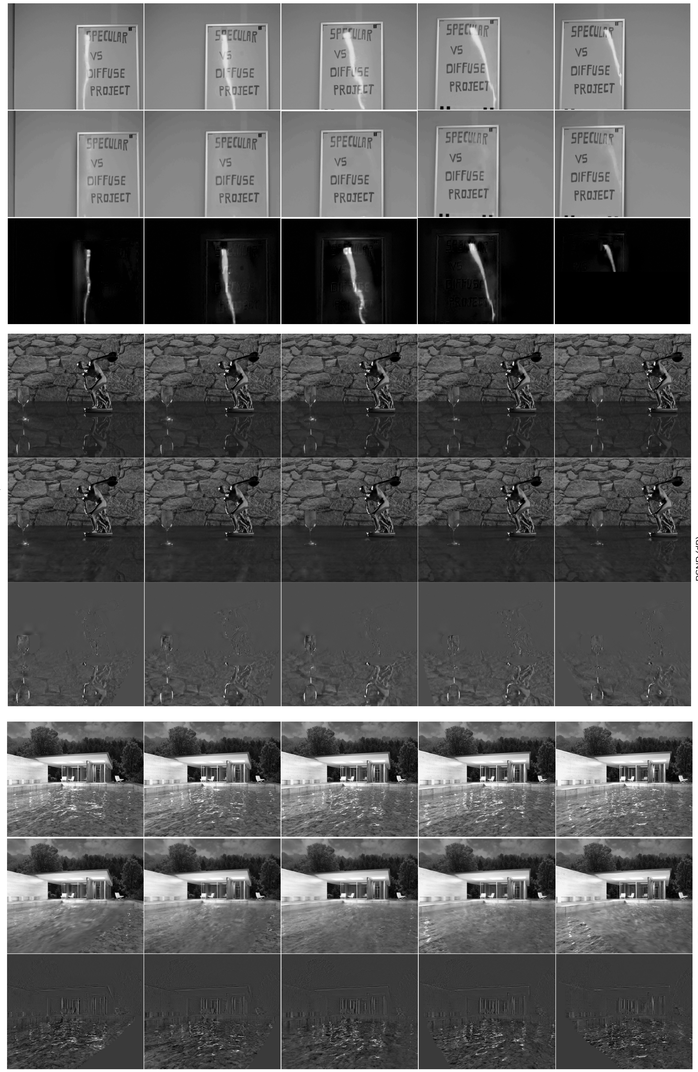
Proposed decomposition using lifting-based 5/3 inter-view transform for diffuse content.

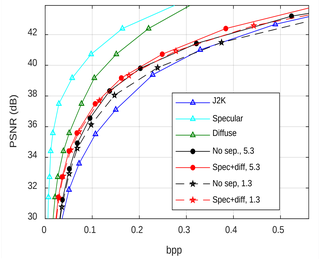


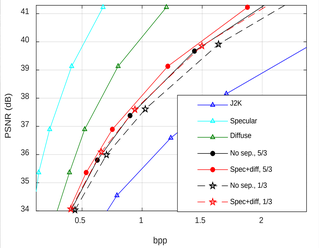
To avoid the complexity of warping between views, we solve the decomposition problem in a registered domain and then warp the solution back to the coordinates of the original views.

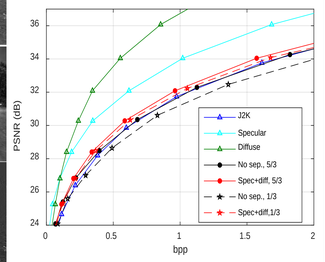


Results









✓ Much of the specular content is removed form input view sequences.

✓ Coding performance is improved in comparison with conventional scalable coding.

Contacts: maryam.haghighat, reji.mathew, d.taubman{@unsw.edu.au}