LEARNING-BASED TONE MAPPING FOR IMAGE MATCHING

Aakanksha Rana, Giuseppe Valenzise, Frédéric Dufaux**

LTCI, Télécom ParisTech, Université Paris-Saclay

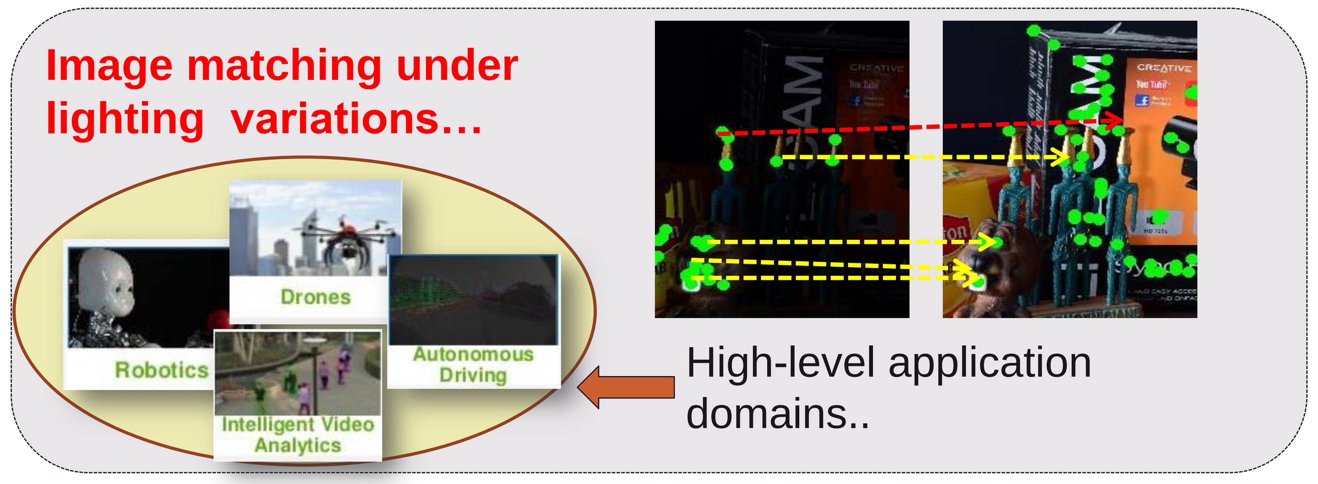
*L2S, CNRS, CentraleSupelec, Université *Paris-Sud*

ABSTRACT

- Propose a new **learning-based tone mapping framework** which
  - relies on a *regression-based* approach to predict locally adaptive parameter-maps,
  - results in tone-mapped images that are optimal for image matching under drastic lighting changes.
- Introduce a mechanism to "generate training samples" using a similarity maximization approach.
- Proposed model
  - evaluated against state-of-the-art TMOs using various descriptor extraction schemes.
  - provides more stable matches in the images undergone drastic lighting variations in the "HDR dataset".

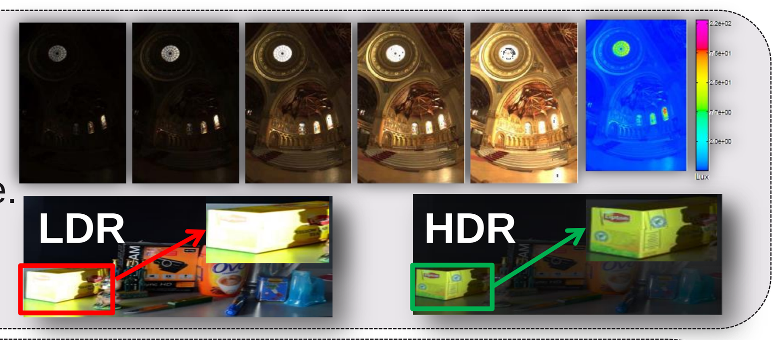
MOTIVATION

Image matching under lighting variations…



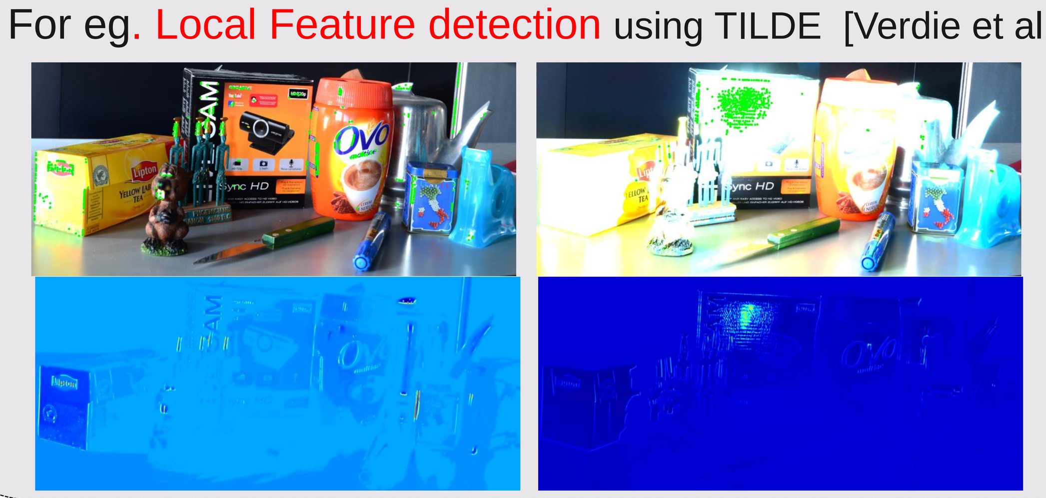
High-level application domains..

HDR :
- Proportional to physical luminance (cd/m^2) of scene.
- Real-Values formats



HDR Linear Values:
- Suboptimal for existing LDR optimal pipelines.

For eg. Local Feature detection using TILDE [Verdie et al.,2015]

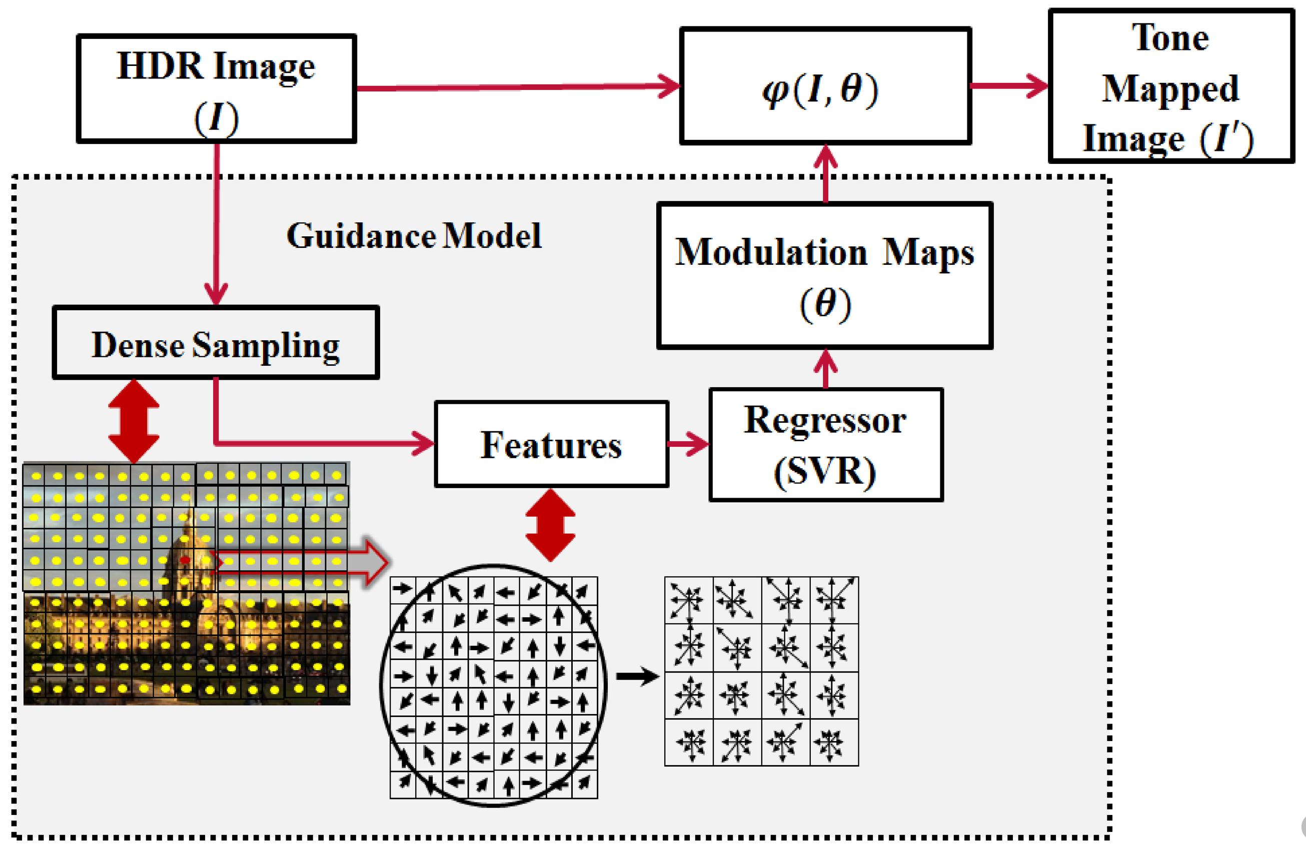


- HDR → **Tone Mapped** HDR for existing LDR friendly algorithms.
- Optimal TMO has been studied in case of keypoint detection [Rana et al 2016]
- What about the feature description ?

Objective:

Optimize the TMOs for
1. invariance to the lighting variations such as day/night changes.
2. Stable matching of descriptors.
3. For increase mean Average precision score.

PROPOSED FRAMEWORK



TMO Framework
- Tone Mapping $I' = \varphi(I,\theta)$ (1), *where* $\theta(x) = \{\theta_1,\ \theta_2\ \dots\}$.
- We demonstrate our model for Bilateral filtering based tone mapping where: $\varphi = I/L$, an Illumination normalization model and L is estimated luminance using Bilateral filter $L = \frac{1}{N}(\sum_{y\in S} G_{\theta_1}(\|x-y\|)\cdot G_{\sigma^r}(\| I_x - I_y \|)\cdot I_y)$. (2) where N is the normalization term.

Generation of Training Set

Selection of training samples
- Identify the 'key' locations in each scene using DoG.
  - Check iteratively for each location if it is detected in majority images undergone lighting variations.
- Randomly select of a keypoint sample location
- Extract SIFT feature defined by gradient orientation given [Dong. et al[15]] as

$$h(\Theta|p)[\mathrm{x}] = \int \mathcal{G}_\delta(\Theta - \angle\nabla p(y))\mathcal{G}_{\hat\sigma}(y-x)\|\nabla p(y)\|d(y)$$

Maximization Objective :

$$\mathcal{F}(\boldsymbol{\theta}) = \frac{1}{K}\sum_{\{i,j\}\in P}\Phi(\boldsymbol{h}_i(\boldsymbol{\theta}), \boldsymbol{h}_j(\boldsymbol{\theta})).$$

where

$$\Phi(\boldsymbol{h}_i, \boldsymbol{h}_j) = \log(1+\exp(\epsilon - \boldsymbol{h}_i^T\boldsymbol{h}_j)$$

SGD based Optimization :

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \gamma_t \cdot \nabla\Phi_{\{i,j\}^t}(\boldsymbol{\theta}_t), \qquad (5)$$

$$\nabla\Phi_{\{i,j\}}(\boldsymbol{\theta}) = \left\{\frac{\partial\Phi}{\partial\mathcal{R}_i}\cdot\frac{\partial\mathcal{R}}{\partial\varphi_i}\cdot\frac{\partial\varphi_i}{\partial\boldsymbol{\theta}}, \frac{\partial\Phi}{\partial\mathcal{R}_j}\cdot\frac{\partial\mathcal{R}}{\partial\varphi_j}\cdot\frac{\partial\varphi_j}{\partial\boldsymbol{\theta}}\right\} \qquad (7)$$

Require: a scene S with N images and the set of possible image pairs P drawn from S.
K:= number of image pairs in P.
$epochs$:= number of passes over the set P.
Ensure:
for $iters = 1 : epochs$ **do**
Shuffle the order of n pairs in P
for $pair = 1 : K$ **do**
Compute $\nabla\Phi_{pair}$ (as in Eq. (7)).
Update $\boldsymbol{\theta}$ (as in Eq. (5)).
end for
end for

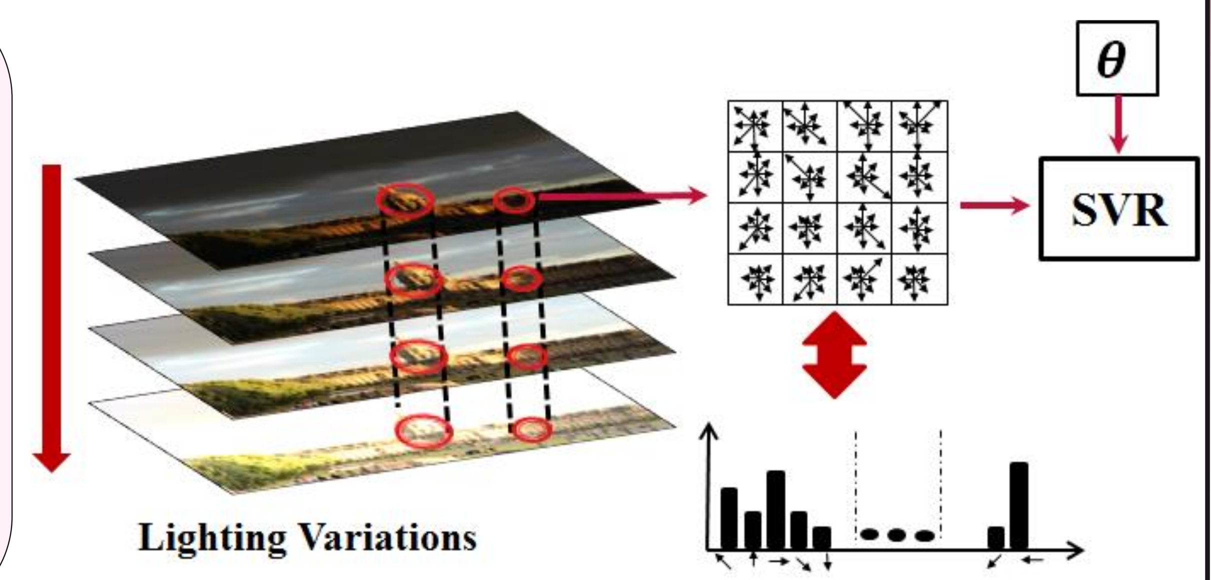
Learning the Prediction Model

- Densly sample the key locations and extract SIFT feature.
- Feed the SVR model with the features and corresponding ground truth and solve the following minimization problem:

$$\min_{\omega,b,\xi,\xi^*} \frac{1}{2}\|\omega^2\| + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)$$

subject to:

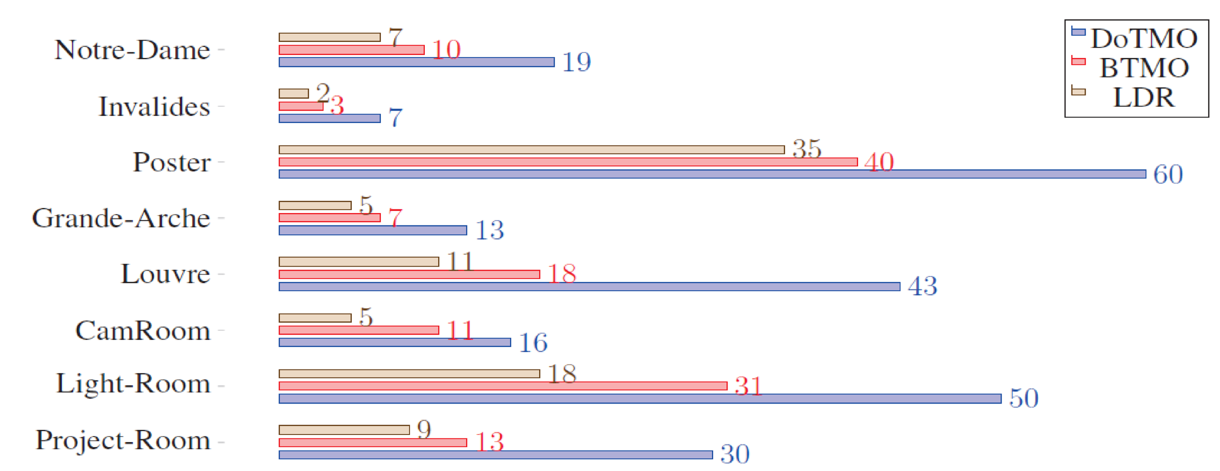
$$\theta_{k(i)} - (\omega^T\psi(f_i) + b) \leq \chi + \xi_i,$$
$$(\omega^T\psi(f_i) + b) - \theta_{k(i)} \leq \chi + \xi_i^*,$$
$$\xi_i, \xi_i^* \geq 0, i = 1..n$$

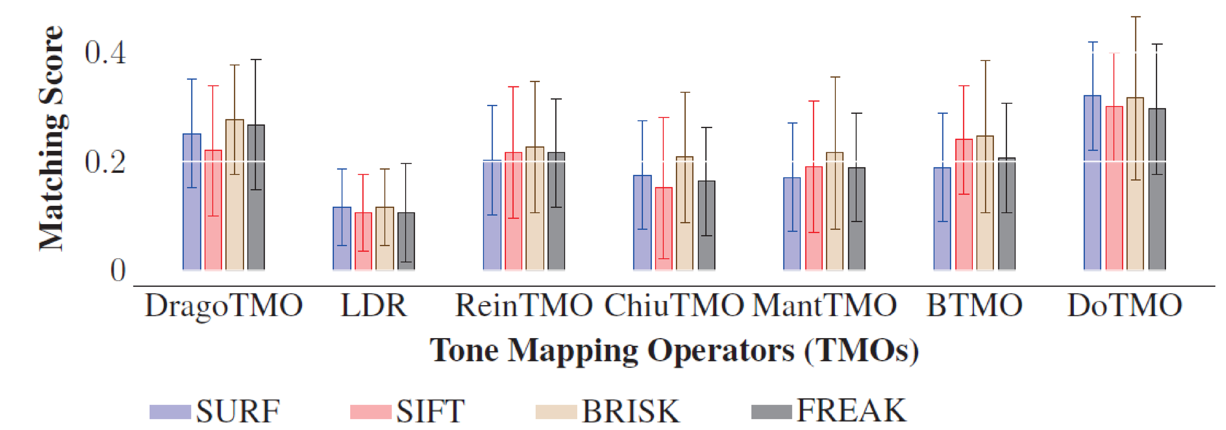


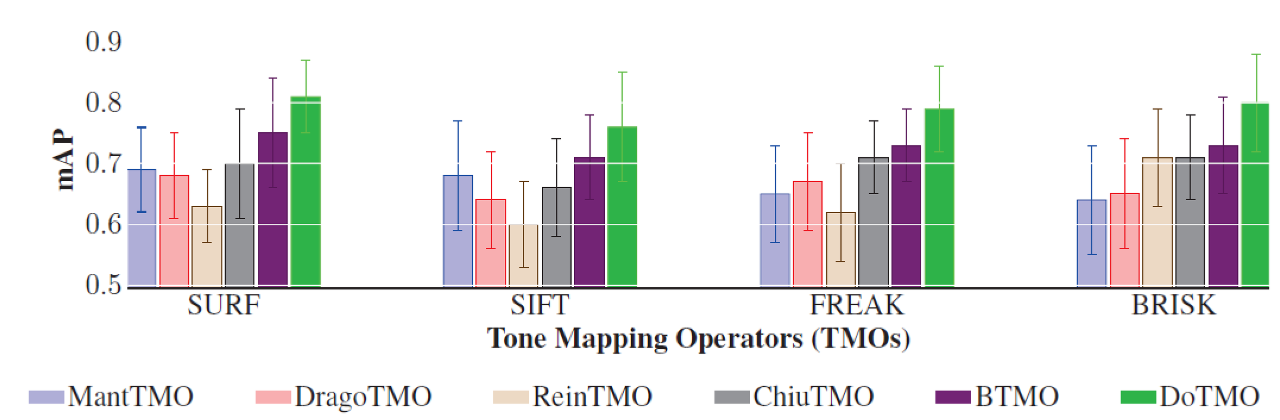
RESULTS & CONCLUSIONS

Experimental Setup
- ✓ HDR dataset with 8 scenes with 6-7 lighting variations.

- ✓ State of the art TMOs: ChiuTMO, DragoTMO, ReinhardTMO, MantiukTMO
- ✓ Descriptor extraction schemes: SURF, SIFT, FREAK, BRISK.
- ✓ Epsilon-SVR used with RBF kernel and regularization cost and epsilon values tuned by 5-fold cross validation from the range $[2^{-5}, 2^{15}]$ and $[2^{-5}, 2^{15}]$.







Day/Night matching using SURF.
Row I: 2 Scene Invalides scene.
Row II: feature matching using our proposed DoTMO (11 correct and 3 incor-rect matches).
Row III: using Reinhard TMO (3 correct and 11 incorrect matches).
Row IV: using MantiukTMO (4 incorrect and 3 correct matches).