Unifying Registration based Tracking: A Case Study with Structural Similarity

Abhineet Singh Mennatullah Siam Martin Jagersand
University of Alberta
asingh1,mennatul,mj7@ualberta.ca

This supplementary presents detailed derivations for the expressions given in the paper for first and second order derivatives of SSIM. Additional results excluded from the main paper due to space constraints are also presented. These include results for ILMs and LSCV. In addition, the results comparing different AMs presented in the paper are reorganized here to compare different SMs with each other for each AM instead. Finally, results for reinitialization tests similar to [7] are also included to be used as an additional metric to evaluate a tracker's robustness. Here the tracker is reinitialized after skipping 5 frames every time its E_{AL} exceeds 20 and the number of such failures is counted. This metric is referred to as the failure rate or **FR**.

1. Derivations for SSIM gradients

1.1. SSIM Jacobian

For clarity and brevity in the subsequent expressions, SSIM is expressed in a simplified form as:

$$f_{ssim} = \frac{ab}{cd} \tag{1}$$

with $a = 2\mu_t\mu_0 + C_1$, $b = 2\sigma_{t0} + C_2$, $c = \mu_t^2 + \mu_0^2 + C_1$ and $d = \sigma_t^2 + \sigma_0^2 + C_2$. We further let $\bar{\mathbf{I}}_\mathbf{t}$ refer to a mean normalized version of $\mathbf{I_t}$ so that $\bar{\mathbf{I}}_\mathbf{t} = \mathbf{I_t} - \mu_t$ and $\sum_{i=1}^{N} \bar{\mathbf{I}}_\mathbf{t}(\mathbf{x_t^i}) = 0$. Differentiating Eq. 1 w.r.t. $\mathbf{I_t}$, we get:

$$\frac{\partial f_{ssim}}{\partial \mathbf{I_t}} = \frac{1}{cd}\left[(a'b + b'a) - f_{ssim}(c'd + d'c)\right] \tag{2}$$

with

$$a' = \frac{\partial a}{\partial \mathbf{I_t}} = \frac{2\mu_0}{N}\frac{\partial \sum_{i=1}^{N} \mathbf{I_t}(\mathbf{x_t^i})}{\partial \mathbf{I_t}} = \frac{2\mu_0}{N}\mathbf{1_N} \tag{3}$$

$$b' = \frac{\partial b}{\partial \mathbf{I_t}} = \frac{2}{N-1}\frac{\partial \sum_{i=1}^{N} \bar{\mathbf{I}}_\mathbf{t}(\mathbf{x_t^i})\bar{\mathbf{I}}_\mathbf{0}(\mathbf{x_0^i})}{\partial \mathbf{I_t}} = \frac{2\bar{\mathbf{I}}_\mathbf{0}}{N-1} \tag{4}$$

$$c' = \frac{\partial c}{\partial \mathbf{I_t}} = \frac{2\mu_t}{N}\frac{\partial \sum_{i=1}^{N} \mathbf{I_t}(\mathbf{x_t^i})}{\partial \mathbf{I_t}} = \frac{2\mu_t}{N}\mathbf{1_N} \tag{5}$$

$$d' = \frac{\partial d}{\partial \mathbf{I_t}} = \frac{1}{N-1}\frac{\partial \sum_{i=1}^{N} (\bar{\mathbf{I}}_\mathbf{t}(\mathbf{x_t^i}))^2}{\partial \mathbf{I_t}} = \frac{2\bar{\mathbf{I}}_\mathbf{t}}{N-1} \tag{6}$$

where $\mathbf{1_N}$ denotes a $1\times N$ vector of all ones. The last equality in Eq. 4 follows since $\forall j \in \{1..N\}$, $\frac{\partial \sum_{i=1}^{N} \bar{\mathbf{I}}_\mathbf{t}(\mathbf{x_t^i})\bar{\mathbf{I}}_\mathbf{0}(\mathbf{x_0^i})}{\partial \mathbf{I_t}(\mathbf{x_t^j})} = \bar{\mathbf{I}}_\mathbf{0}(\mathbf{x_0^j}) - \frac{1}{N}\sum_{i=1}^{N} \bar{\mathbf{I}}_\mathbf{0}(\mathbf{x_0^i}) = \bar{\mathbf{I}}_\mathbf{0}(\mathbf{x_0^j})$. Similar reasoning holds for Eq. 6 too. Substituting Eqs. 3 - 6 in Eq. 2 gives:

$$\frac{\partial f_{ssim}}{\partial \mathbf{I_t}} = \frac{2}{cd}\left[\left(\frac{\mu_0 b}{N}\mathbf{1_N} + \frac{a\bar{\mathbf{I}}_\mathbf{0}}{N-1}\right) - f_{ssim}\left(\frac{\mu_t d}{N}\mathbf{1_N} + \frac{c\bar{\mathbf{I}}_\mathbf{t}}{N-1}\right)\right] \tag{7}$$

1.2. SSIM Hessian

Referring to f_{ssim} as f and $\frac{\partial f_{ssim}}{\partial \mathbf{I_t}}$ as $\mathbf{f}'$ for breviy and letting $\mathbf{A} = \frac{b\mu_0 - fd\mu_t}{N}\mathbf{1_N}$, $\mathbf{B} = \frac{a\bar{\mathbf{I}}_\mathbf{0} - fc\bar{\mathbf{I}}_\mathbf{t}}{N-1}$ and $D = \frac{cd}{2}$, Eq. 7 can be rewritten as:

$$\mathbf{f}' = \frac{\mathbf{A} + \mathbf{B}}{D} \tag{8}$$

Differentiating Eq. 8 w.r.t. $\mathbf{I_t}$, we get:

$$\frac{\partial^2 f}{\partial \mathbf{I_t}^2} = \frac{1}{D}\left((\mathbf{A}' + \mathbf{B}') - \mathbf{f}'^T\mathbf{D}'\right) \tag{9}$$

with

$$\mathbf{A}' = \frac{\partial \mathbf{A}}{\partial \mathbf{I_t}} = S_N\left(\frac{1}{N}\left[\mu_0 b' - \mu_t(d\mathbf{f}' + fd') - \frac{fd}{N}\mathbf{1_N}\right]\right) \tag{10}$$

$$\mathbf{B}' = \frac{\partial \mathbf{B}}{\partial \mathbf{I_t}} = \frac{1}{N-1}\left(\bar{\mathbf{I}}_\mathbf{0}a' - \bar{\mathbf{I}}_\mathbf{t}(c\mathbf{f}' + fc') - \mathrm{diag}\left(fc\mathbf{1_N}\right)\right) \tag{11}$$

$$\mathbf{D}' = \frac{\partial D}{\partial \mathbf{I_t}} = \frac{dc' + cd'}{2} = \frac{1}{2}\left(\frac{\mu_t d}{N}\mathbf{1_N} + \frac{c\bar{\mathbf{I}}_\mathbf{t}}{N-1}\right) \tag{12}$$

where $S_n(\mathbf{K})$ denotes an $n\times k$ matrix formed by stacking the $1\times k$ vector $\mathbf{K}$ into rows while $\mathrm{diag}(\mathbf{K})$ denotes a $k\times k$

diagonal matrix with $\mathbf{K}$ as the principal diagonal. Substituting Eqs. 10, 11, 12 in Eq. 9 and simplifying gives:

$$\frac{\partial^2 f_{ssim}}{\partial {\mathbf{I_t}}^2} = \frac{2}{cd}\left[\frac{1}{N} S_N \left(\frac{4}{N-1}\left(\mu_0 \bar{\mathbf{I}}_\mathbf{0} - \mu_t f \bar{\mathbf{I}}_\mathbf{t}\right) - \frac{3\mu_t d}{2}\mathbf{f}' - \frac{fd}{N}\right) - \frac{c}{N-1}\left(\frac{3}{2}{\mathbf{f}'}^T \bar{\mathbf{I}}_\mathbf{t} + f\mathbb{I}\right)\right] \quad (13)$$

where $\mathbb{I}$ is an $N \times N$ identity matrix.

2. Results and Analysis

2.1. Search Methods

The plots comparing different SMs with all AMs in terms of both SR and FR are given in Figs. 1- 3. For additional clarity, the SR plots have been divided into two parts so the results for each AM comprise three plots with the third one showing the FR. Following are some notable observations from these results:

- PFFC usually performs best followed by RKLT. The only exceptions are CCRE and SCV both of which perform best with NNIC.
- RKLT performs best relative to other SMs with pixel level AMs like SSD and SPSS even though these are the simplest ones. This is probably because the low DOF trackers running on small sub patches benefit from an AM that is less invariant to small changes in the patch's appearance.
- RANSAC peforms quite similarly to RKLT except for smaller thresholds where its stochastic nature causes it be more jittery and so less precise. Also, it is usually the best performer among stochastic SMs.
- Both NN and NNIC are usually the worst performing SMs among all stochastic and hybrid SMs respectively.
- ESM fails to outperform FCLK/FALK for any AM except SCV and even there it is probably the tendency of SCV to favor inverse models that is the main reason for ESM being better. This hypothesis is given more weight by the fact that ESM is just as much worse with RSCV as it is better with SCV. This fact too emerges in contradiction to the theoretical analysis in [3] where ESM was shown to have second order convergence and so should be better than first order methods like FCLK and FALK. It might be argued that the extended version of ESM [4, 11] used here might not possess the characteristics of the formulation described in [3] but experiments were conducted with that exact formulation too and it was not found to perform significantly different from the one used here.
- The gain in performance with NNIC compared to ICLK is more marked with respect to FR than SR, with the former having been reduced by around half with all the AMs. The same holds true to a lesser extent with PFFC and RKLT when compared with FCLK. This is probably because the main advantage of providing a better starting point for GD search is to make the tracker more robust by avoiding failures in scenarios like fast motion where the object's location in the last frame lies outside the region of convergence.
- IALK fares even worse against ICLK with robust AMs than with the SSD-like AMs. This is consistent with the fact that the extension of GN method with $\hat{\mathbf{H}}_{self}$ does not make as much sense for additive SMs as compositional ones [5].
- The performance improvement provided by PFFC and NNIC over FCLK and ICLK respectively is more strongly marked with robust AMs than with SSD-like models especially in terms of SR. In fact, PFFC with NCC proves to be the best tracker tested in this study and provides, to the best of our knowledge, a new state of the art in high DOF tracking.
- CCRE seems to favor inverse models as both ICLK and NN fare better against FCLK and PF respectively than with most other AMs. This trend becomes even more noticeable when comparing NNIC with PFFC. This is somewhat surprising since this is also the worst performing AM and so presumably harder to optimize, which would be expected to work better with the more sophisticated forward models.

2.2. Appearance Models

The plots comparing different AMs for each SM are given in Figs. 4 - 6. As in the previous section, the results for each SM are again divided into two parts for clarity - SSD like AMs and robust AMs. The accuracy and robustness plots comparing different AMs are shown in Fig. **??**. Some interesting points to be noted are listed below:

- LSCV, in spite of being the most sophisticated and computationally expensive variant of SCV, usually performs worse than both the original formulation as well as RSCV.
- CCRE performs worst among all AMs for all SMs except NN and NNIC.
- Though SSIM fails to outperform NCC on average, it is usually very close and can be regarded as a similar performer on average which also makes it one of the best among all tested AMs.
- SPSS and SSD perform almost identically with GD based SMs and only slightly better than CCRE which is to be expected as both are simple pixel wise measures.

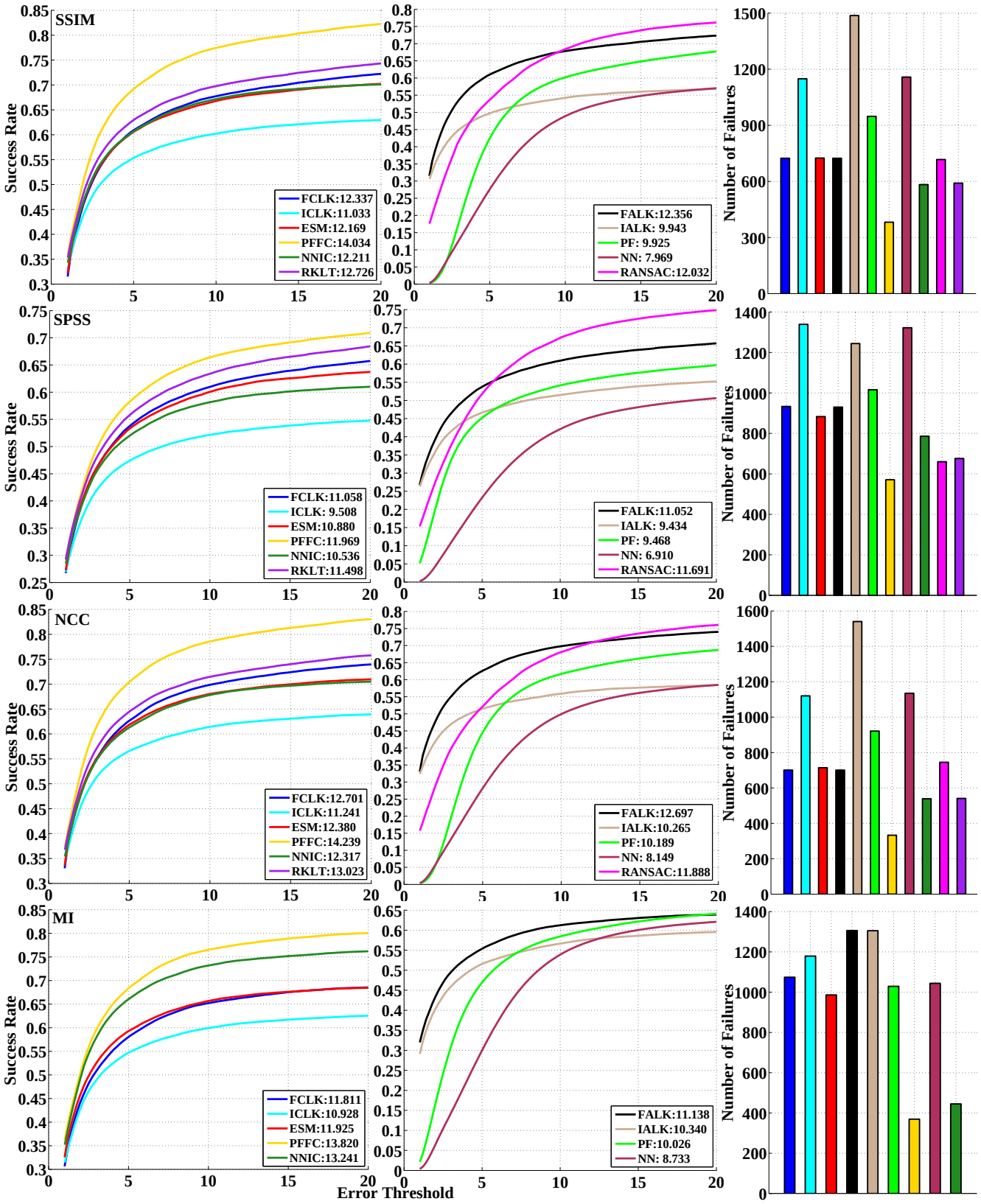

Figure 1: Success rates for SMs using SSIM, SPSS, NCC and MI with Homography. Note that the range of y axis varies between plots to maximize visibility in each. Best viewed on a high resolution screen.

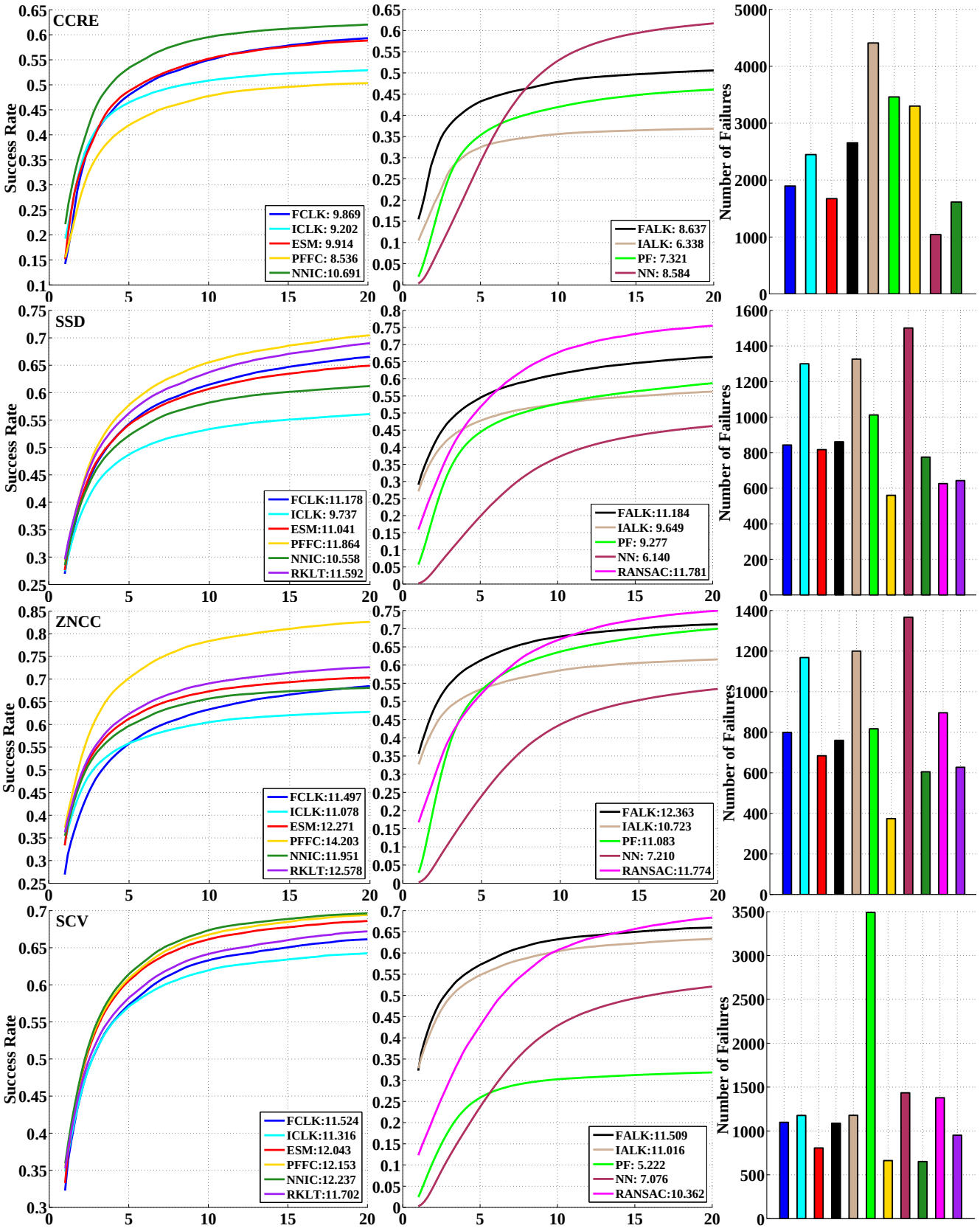

Figure 2: Success rates for SMs using CCRE, SSD, ZNCC and SCV with Homography. Note that the range of y axis varies between plots to maximize visibility in each. Best viewed on a high resolution screen.

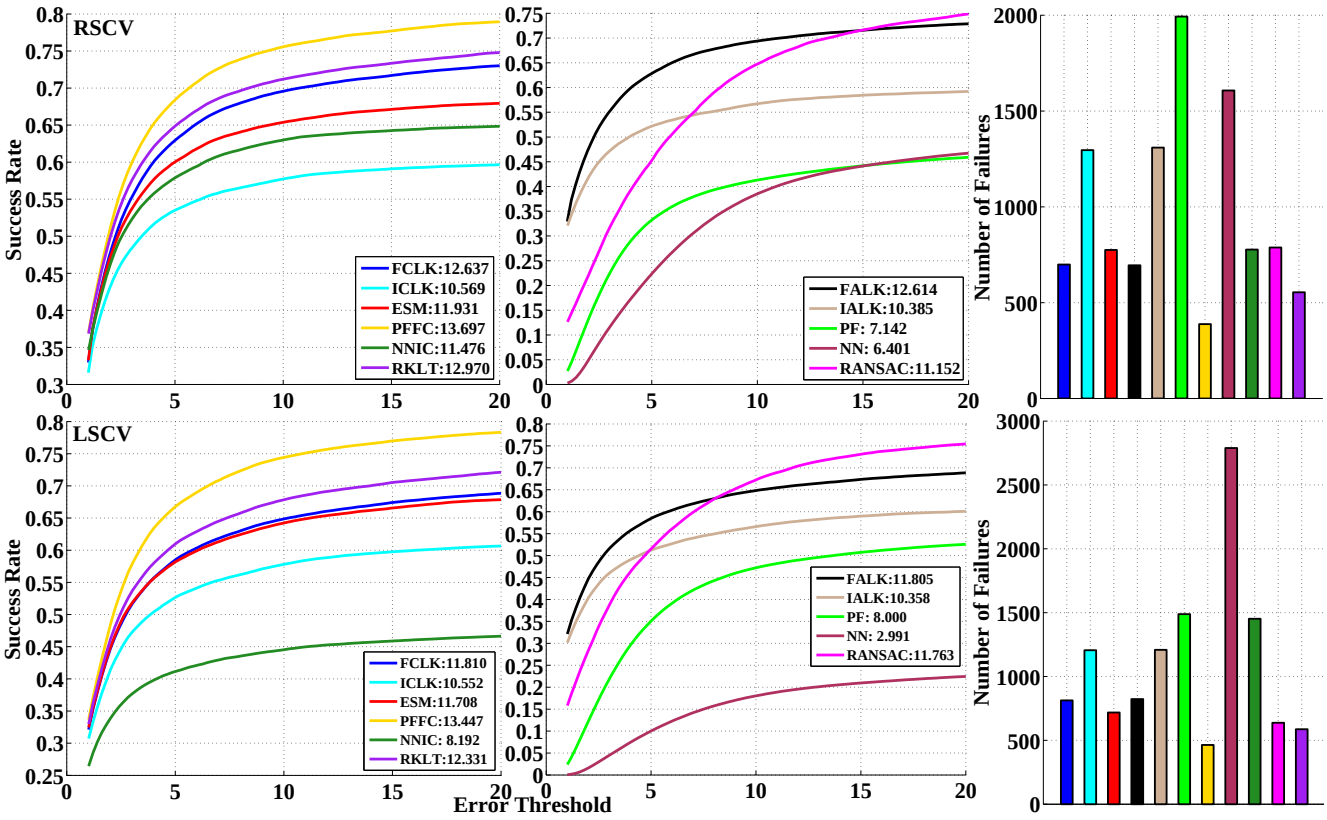

Figure 3: Success rates for SMs using RSCV and LSCV with Homography. Note that the range of y axis varies between plots to maximize visibility in each. Best viewed on a high resolution screen.

- The accuracies of different AMs are more consistent with NN based SMs than with GD or RANSAC based methods.

- FCLK, FALK and ESM perform best with NCC followed by RSCV, SSIM and ZNCC. The superiority of NCC is worth noting considering that it is a fairly simple and long known image similarity metric but still manages to outperform newer, more sophisticated and computationally expensive AMs like MI, CCRE and NGF. The near identical performance of NCC and SSIM is probably to be expected though, as their functional forms too are fairly similar.

- A clear role reversal can be observed between SCV and RSCV in their relative performance with the forward and inverse SMs - SCV outperforms RSCV with the latter by almost the same amount as it is outperformed by RSCV with the former. Also, the two AMs perform almost identically with ESM that combines both the forward and inverse methods - SCV is slightly better in SR and RSCV in FR. This can be explained by the fact that SCV replaces $\mathbf{I_0}$ with the expected patch $\mathbb{E}(\mathbf{I_t}|\mathbf{I_0})$ [8] while RSCV replaces $\mathbf{I_t}$ with $\mathbb{E}(\mathbf{I_0}|\mathbf{I_t})$ [6].

- Though ZNCC does perform similarly to NCC, the latter seems to be slightly better overall especially with the forward SMs. Assuming that minimizing the SSD between normalized patches is equivalent to maximizing the dot product between them [10], the reverse would be expected due to the wider convergence region of ZNCC so it would appear that this equivalence does not hold in practice.

- MI and CCRE are the best performing AMs with NN. This indicates that their relatively poor performance with GD based SMs is more due to its narrow basin of convergence rather than any inherent shortcoming in the similarity measures themselves.

- When used with NNIC, however, CCRE reverts to being one of the worst performing AMs. It seems that the radius of convergence of CCRE is too narrow for ICLK to converge even when a better starting point is provided by NN. MI does not appear to suffer from this drawback and so ends up being the best performing AM with NNIC and by a significant margin too. This is consistent with the better performance of MI with ICLK - apparently its radius of convergence is

large enough for the coarse location provided by NN to lie within it more often than not.

- SCV outperforms RSCV with both NN and NNIC which is consistent with the observation made with GD SMs that it works well with SMs that search I_0 for the optimal warp.

- RSCV outperforms SCV with PF, thus continuing the trend of performing well with forward SMs just as SCV had done with NN, thus reaffirming that the two should respectively be used with forward and inverse SMs.

- ZNCC is the best performing AM with PF followed by NCC and MI. MI continues to perform well with stochastic SMs to add weight to the hypothesis that its relatively poor performance with GD based SMs is mainly due to its narrow convergence region and not any inherent shortcoming as a similarity measure. ZNCC may partially owe its good performance to the fine tuned likelihood function and associated constants that were taken from [9].

- The unexpectedly poor performance of CCRE in particular might be due to insufficiently optimized likelihood constants though the best ones that could be obtained, given the time constraints, were used.

- CCRE is the worst performer on PFFC which is consistent with its poor performance with both PF and FCLK. Since the PF layer does not perform well, its results probably provide even worse starting location for FCLK than that from the last frame. As a result CCRE actually performs worse with PFFC than FCLK. This is one of the main downsides of cascade tracking - poor quality results from the first layer can end up causing the second layer to perform even worse than it would by itself.

2.3. Illumination Models

Results comparing the three ILMs - GB [1, 2], PGB [13, 14, 15, 12] - and RBF [13] are presented in Fig. 7. SSD is the only AM in MTF that currently supports ILMs so testing is limited to this AM. As the simplest AM, SSD is also likely to be more sensitive the addition of ILMs and so will demonstrate their impact better. NCC is also shown for comparison as it too provides illumination invariance like the ILMs. PGB was run by dividing the tracked patch into 3×3 grid of sub patches so that $A = 10$ photometric parameters were optimized along with the 8 parameters of homography. Similarly, RBF was run using a 3×3 grid control points.

Following observations can be made there:

- GB does improve performance over SSD with all three SMs but fails to match NCC with any of them. This is rather surprising as NCC, like GB, supposedly provides invariance only against affine illumination changes so the latter should be able to handle these just as well. The 2 extra parameters that need to be estimated with GB might be the reason for this.

- RBF improves further over GB, at least in terms of SR, and does manage to perform as well as NCC with both ICLK and ESM. In terms of FR, however, it outperforms GB only with ESM and has higher FR than both GB and SSD with the other two SMs. This is probably because the 10 extra parameters that need to be estimated make the search more likely to get stuck in local maxima thus causing the tracker to fail.

- The significantly lower FR of RBF with ESM finally lends some credence to the supposedly superior convergence properties of this SM. In fact, ESM in general seems to handle higher DOF ILMs better than FCLK which in turn improves over ICLK. This seems to indicate that the advantage of using a more sophisticated SM becomes more prominent as the dimensionality of the search space increases, thus rendering the search more challenging.

- PGB performs worse than GB in terms of both FR and SR with all of the SMs though its poor performance is most notable with ICLK, where it is even worse than SSD. It seems that the discontinuity between subregions that is inherent to this ILM does not represent realistic lighting changes well and the disadvantageous effect of the extra parameters dominates.

References

[1] A. Bartoli. Direct image registration with gain and bias. *Contributions au recalage dimages et à la reconstruction 3D de scènes rigides et déformables*, page 4, 2006.

[2] A. Bartoli. Groupwise geometric and photometric direct image registration. *PAMI, IEEE Transactions on*, 30(12):2098–2108, 2008.

[3] S. Benhimane and E. Malis. Homography-based 2D Visual Tracking and Servoing. *Int. J. Rob. Res.*, 26(7):661–676, July 2007.

[4] R. Brooks and T. Arbel. Generalizing Inverse Compositional and ESM Image Alignment. *IJCV*, 87(3):191–212, May 2010.

[5] A. Dame. *A unified direct approach for visual servoing and visual tracking using mutual information*. PhD thesis, University of Rennes, 2010.

[6] T. Dick, C. Perez, M. Jagersand, and A. Shademan. Real-time Registration-Based Tracking via Approximate Nearest Neighbour Search. In *Proceedings of Robotics: Science and Systems*, Berlin, Germany, June 2013.

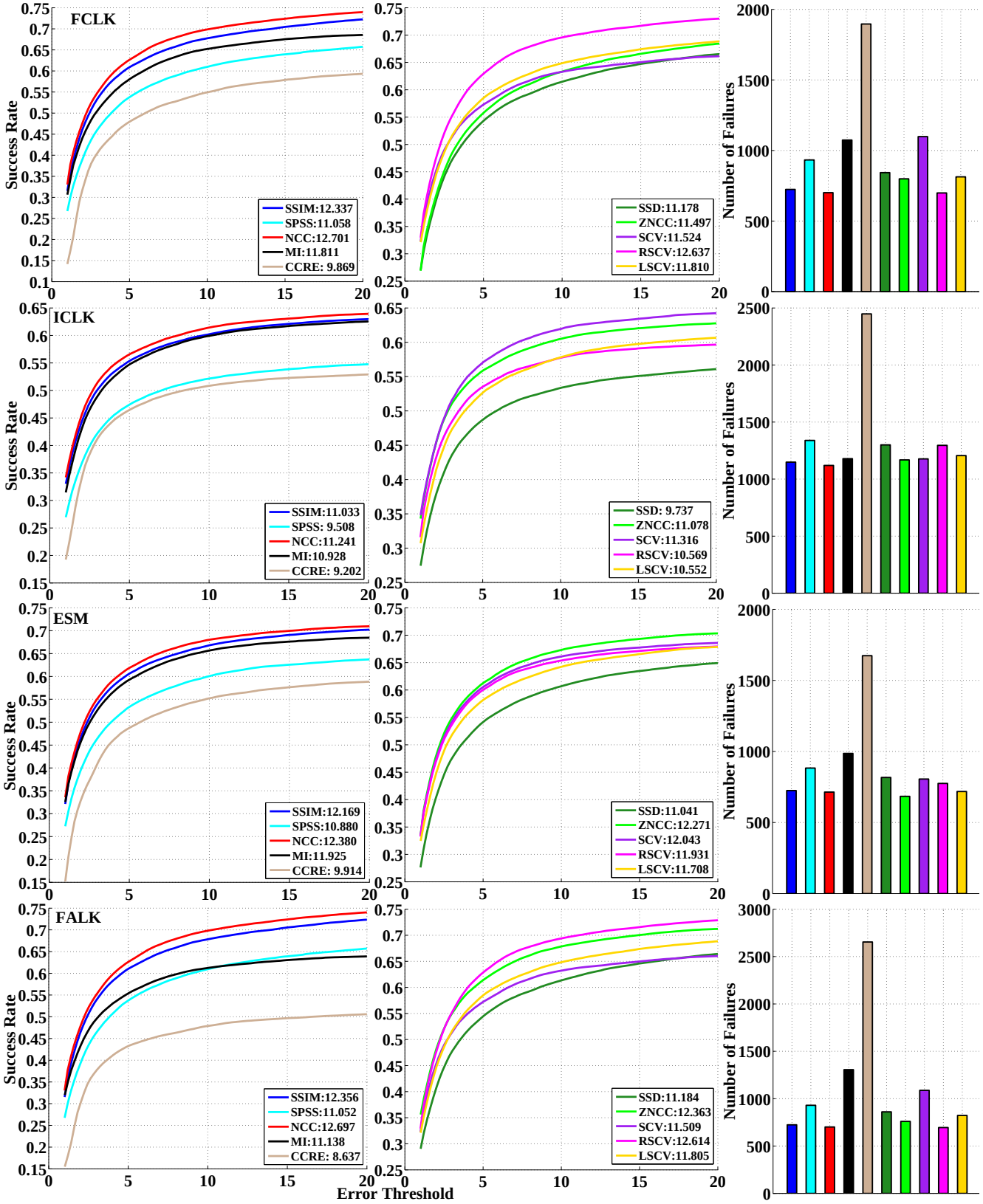

Figure 4: Success rates for AMs FCLK, ICLK ESM and FALK with Homography.

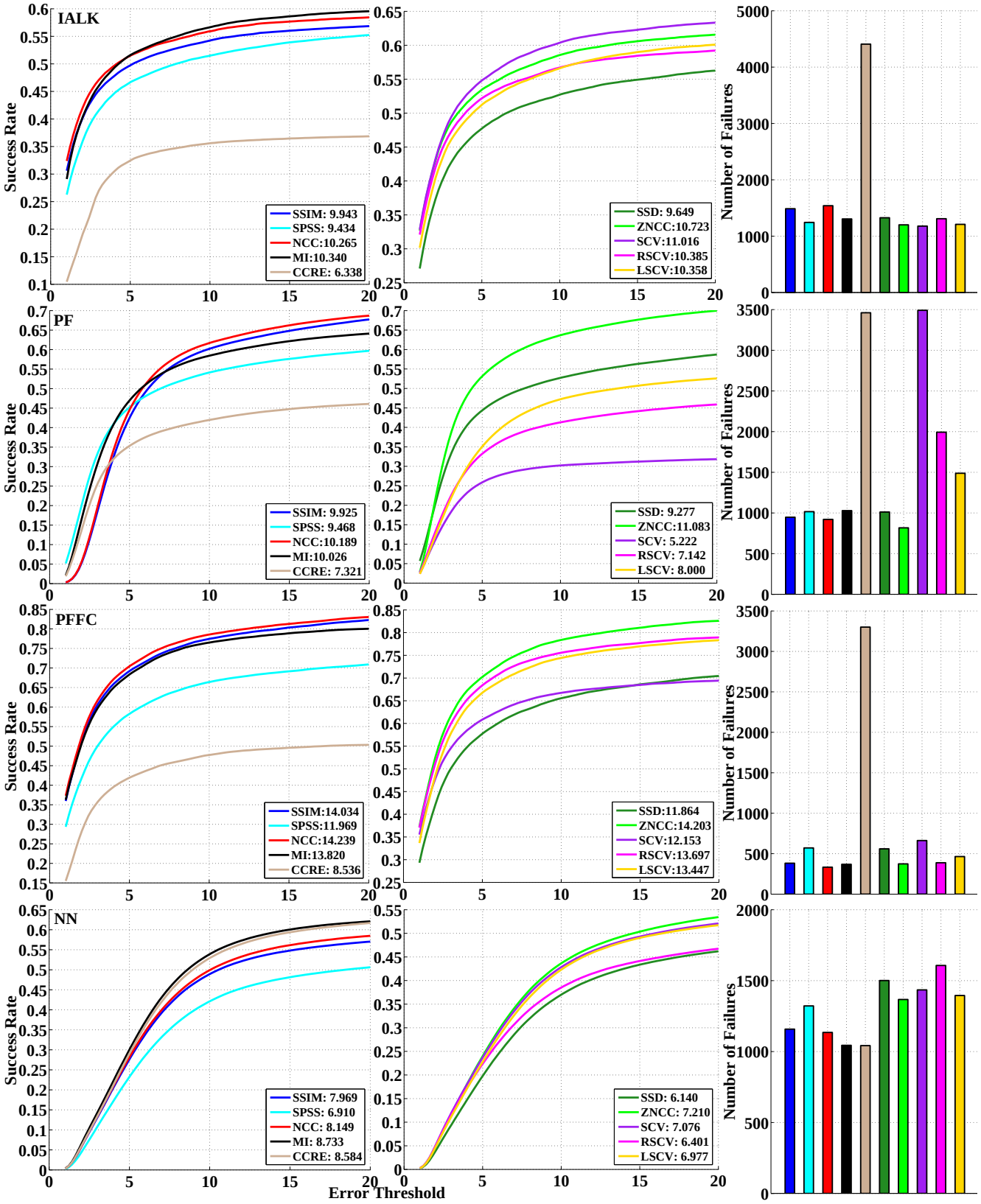

Figure 5: Success rates for AMs using IALK, PF, PFFC and NN with Homography.

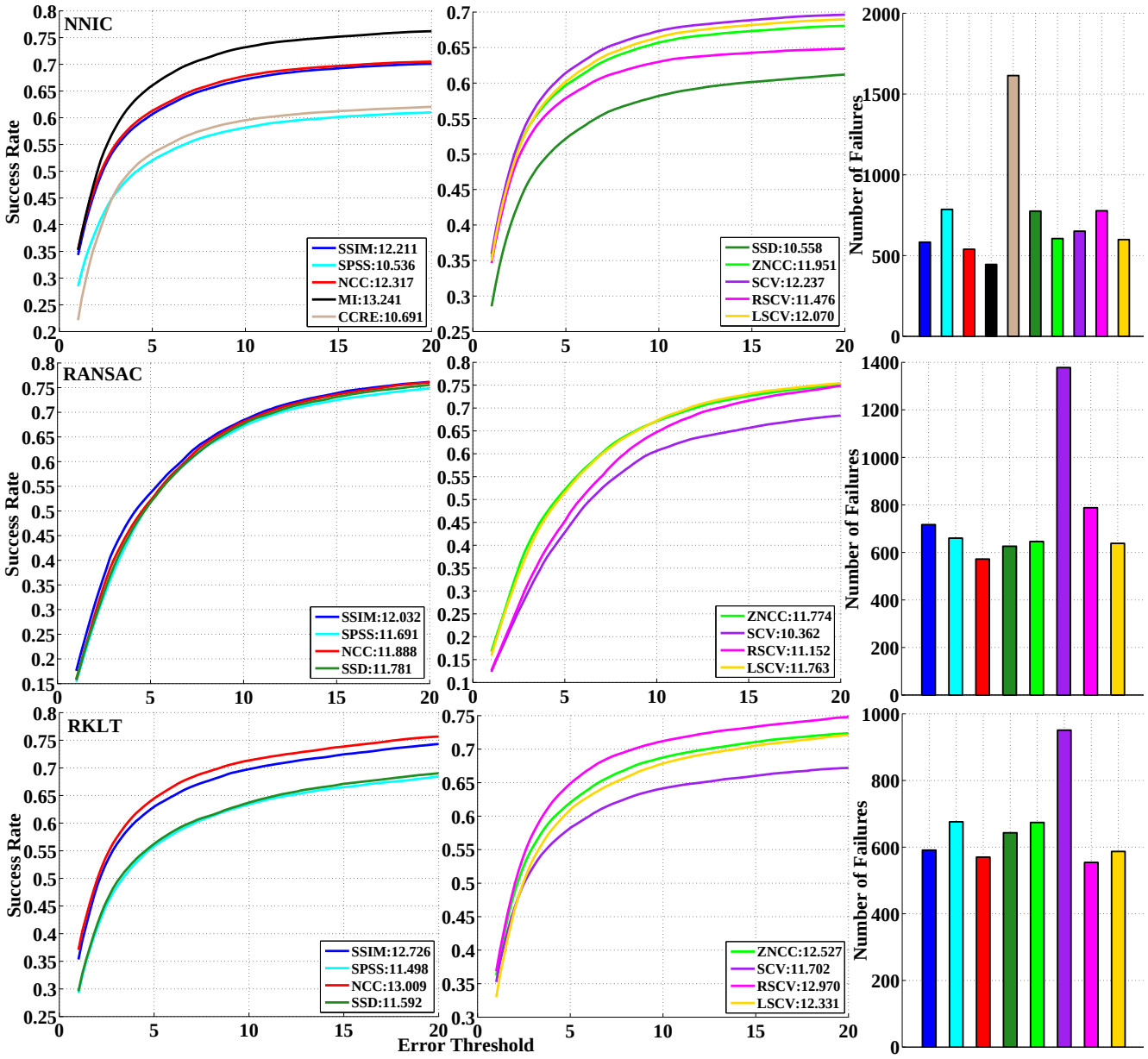

Figure 6: Success rates for AMs using NNIC, RANSAC and RKLT with Homography.

[7] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, et al. *The Visual Object Tracking VOT2016 Challenge Results*, pages 777–823. Springer International Publishing, Cham, 2016.

[8] R. Richa, R. Sznitman, R. Taylor, and G. Hager. Visual tracking using the sum of conditional variance. In *IROS, IEEE/RSJ International Conference on*, pages 2953–2958, Sept 2011.

[9] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang. Incremental Learning for Robust Visual Tracking. *IJCV*, 77(1-3):125–141, May 2008.

[10] L. Ruthotto. Mass-preserving registration of medical images. *German Diploma Thesis (Mathematics), Institute for Computational and Applied Mathematics, University of Münster*, 2010.

[11] G. G. Scandaroli, M. Meilland, and R. Richa. Improving NCC-based Direct Visual Tracking. In *ECCV*, pages 442–455. Springer, 2012.

[12] G. Silveira. Photogeometric Direct Visual Tracking for Central Omnidirectional Cameras. *Journal of Mathematical Imaging and Vision*, 48(72), October 2014.

[13] G. Silveira and E. Malis. Real-time visual tracking under arbitrary illumination changes. In *CVPR. IEEE Conference on*, pages 1–6, 2007.

[14] G. Silveira and E. Malis. Visual servoing from robust direct

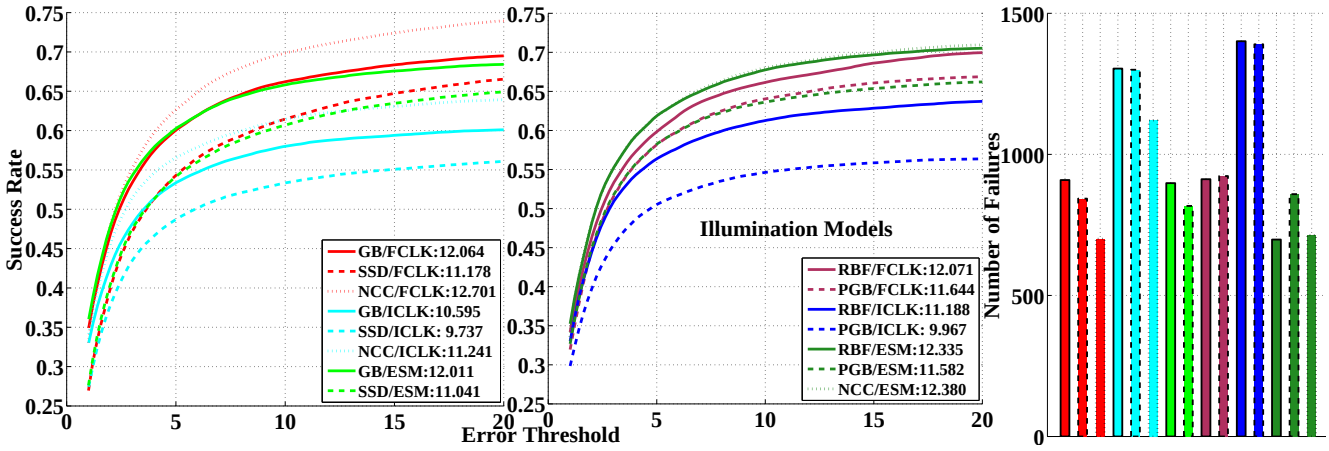

Figure 7: Success rates for ILMs using FCLK, ICLK and ESM with Homography.

color image registration. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5450–5455. IEEE, 2009.

[15] G. Silveira and E. Malis. Unified direct visual tracking of rigid and deformable surfaces under generic illumination changes in grayscale and color images. *International journal of computer vision*, 89(1):84–105, 2010.