ANCC FLOW: ADAPTIVE NORMALIZED CROSS-CORRELATION WITH EVOLVING GUIDANCE AGGREGATION FOR DENSE CORRESPONDENCE ESTIMATION

Seungryong Kim†  Dongbo Min†  Kwanghoon Sohn‡

†School of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea
‡Department of Computer Science and Engineering, Chungnam National University, Daejeon, Korea
E-mail: khsohn@yonsei.ac.kr

ABSTRACT

Adaptive normalized cross-correlation (ANCC) cost function works well between images under photometric distortions, but its heavy computational burden often limits its applications. To overcome this limitation, this paper proposes a robust and efficient computational framework, called ANCC flow, designed for establishing dense correspondences between images under severe photometric variations. We first simplify the weight of ANCC in an asymmetric manner by considering a source image weight only. It is then efficiently computed by applying constant-time edge-aware filters without loss of its matching accuracy. Additionally, to deal with a large discrete label space effectively, which is a challenging issue in flow field estimations, we propose a randomized label space sampling strategy similar to PatchMatch filter (PMF) optimization. The robustness of the asymmetric ANCC and the cost filter is further enhanced through an evolving weight computation, where a flow field computed in a previous iteration is utilized to build current edge-aware weights. Experimental results demonstrate the outstanding performance of ANCC flow in many cases of dense correspondence estimations under severe photometric and geometric variations.

Index Terms— adaptive normalized cross-correlation, Patchmatch filter, dense correspondence, stereo matching

1. INTRODUCTION

In many computer vision and computational photography applications, images captured under different imaging modalities are popularly used to overcome their inherent limitations, such as flash and no-flash images [1], color and dark flash images [2], blurred images [3, 4], and images under different radiometric conditions [5].

To realize these tasks, establishing dense correspondences between image pairs across photometric variations is an essential problem. Conventional methods for estimating depth [6] or optical flow [7, 8] fields, in which input images are acquired in a similar imaging condition, have been dramatically advanced in recent studies. In these approaches, a matching fidelity term is not a critical issue, as they assume that multiple images share a similar visual pattern. Instead, they focus on powerful labeling optimizers, e.g., graph-cut (GC) [9], PatchMatch [10], cost volume filter [11], non-rigid dense correspondence (NRDC) [12], and Patchmatch filter (PMF) [13]. However, for images taken under different modality conditions, they cannot deal with severe photometric variations without suitable matching cost functions or descriptors [14]. In those cases, robust cost functions designed to deal with modality variations are one of the most important issues to yield a reliable matching quality [15]. Unfortunately, conventional gradient-based descriptors such as scale invariant feature transform (SIFT) [16] and DAISY [17], as well as intensity comparison-based binary descriptors such as binary robust independent elementary features (BRIEF) [18], have shown a limited performance in matching under photometric variations [15].

To overcome these limitations, a number of methods have been proposed, and could be categorized as feature-based and area-based approaches. In feature-based approaches, several methods tried to reformulate existing descriptors [19], but they showed limitations similar to existing descriptors. Schechtman and Irani introduced local self-similarity (LSS) descriptor [20], and achieved impressive results in object detection and retrieval. Based on the LSS [20], several methods applied it to multi-modal registration problems [21, 22, 23, 15]. These feature-based approaches have shown satisfactory results in a robust manner, but their discriminative power is limited, leading to difficulty of accurate matches especially on boundary regions.

Among area-based approaches, mutual information (MI)-based cost function is used for a registration of multi-modal medical images [24]. As a pioneering work, the cross-correlation (CC)-based cost function has been popularly used for multi-modal image correspondences, e.g., normalized CC (NCC) [6], adaptive NCC (ANCC) [25], Mahalanobis distance CC (MDCC) [26], and robust selective normalized CC (RSNCC) [27]. Although they provide satisfactory results, when the search space is large, their computational time is also inevitably high [15]. When incorporated with an global energy function and solved by a global optimizer [9], their complexity further dramatically increases.

In this paper, our approach is focusing on the ANCC cost function [25], which has shown satisfactory results on correspondences under photometric variations, but has inherently high computational burdens. We reformulate ANCC [25] in a robust and efficient manner, and further combine it with a cost volume filtering-based optimization [11, 13]. Specifically, our approach approximates ANCC [25] by considering a source guidance weight only, which enables us to apply constant time edge-aware filters (EAF) [28] for a fast computation. To reduce a computational burden for large search spaces, we further employ PMF-like random search strategy. In the optimization procedure, by leveraging evolving guidance weights in computing the cost function and cost volume filter, more reliable

Fig. 1. Comparison of ANCC flow with existing approaches. (a), (b) stereo images, and depth maps with (c) NCC+GC, (c) ANCC+GC, and (d) ANCC flow. ANCC flow runs 60× faster than ANCC+GC while providing a lower error rate (5.11% vs. 12.24%).
and robust flow fields are estimated as the iterations. Unlike other methods, our ANCC flow can be easily extended to overcome geometric variation problems between image pairs. In experiments for image pairs under photometric and even geometric variations, our ANCC flow outperforms conventional feature-based and area-based approaches both quantitatively and qualitatively.

2. THE ANCC COST FUNCTION AND ITS LIMITATIONS

Let us define an image as $f_i: \mathcal{I} \rightarrow \mathbb{R}$ for pixel $i = [x_i, y_i]^T$, where $\mathcal{I} \subset \mathbb{N}^2$ is a discrete image domain. Given a pair of images $f_i$ and $f_j$, a dense correspondence estimation aims to assign each pixel $i$ a label $l_i \in \mathcal{L} = \{l = [u_i, v_i]^T\}$, satisfying that $f_{i+l_i} = f_j$.

Unlike conventional cost measures [6], an adaptive normalized cross-correlation (ANCC) cost function deals with photometric variations between multiple images effectively, by leveraging its edge-aware subtraction and normalization [25]. Given a pixel $i$ and its label candidate $l$, the ANCC cost function $\Phi(i, l)$ is defined between two patches $\mathcal{F}_i$ for pixel $i$ of source image $f'$ and $\mathcal{F}_j$ for corresponding pixel $j$ (where $j = i - l$) of target image $f^j$ as

$$\Phi(i, l) = \frac{\sum_{i',j'} \omega_{i,i'} \omega_{j,j'} (f_{i'} - G_i) (f_{j'} - G_j)}{\sqrt{\sum_{i'} \omega_{i,i'} (f_{i'} - G_i)^2} \sqrt{\sum_{j'} \omega_{j,j'} (f_{j'} - G_j)^2}}. \quad (1)$$

where $i' \in \mathcal{F}_i$ and $j' \in \mathcal{F}_j$, and weighted averages on $\mathcal{F}_i$ and $\mathcal{F}_j$ are defined such that $G_i = \sum_{i'} \omega_{i,i'} f_{i'}$ and $G_j = \sum_{j'} \omega_{j,j'} f_{j'}$. $\omega_{i,i'}$ and $\omega_{j,j'}$ is the normalized adaptive weight of a support pixel $i'$ defined on the image. Based on the cost $\Phi(i, l)$ for all pixel $i$ and candidate $l \in \mathcal{L}$, the final flow fields are generally inferred by minimizing the energy function with a global optimizer, e.g. graph-cut (GC) [9].

Although it provides outstanding performances, it has inherent limitations to be applied to a general dense correspondence scenario. Due to its computation of edge-aware weights $\omega_{i,i'}$ and $\omega_{j,j'}$, a computational time dramatically increases as a large size of support window $|\mathcal{F}|$ and search space $|\mathcal{L}|$ [25]. Furthermore, the edge-aware weight defined on a color image only leads to texture copy problems from the color image [29] and building non-reliable cost volumes, which limits the matching performance.

3. THE ANCC FLOW

3.1. Overview of Our Approach

Our ANCC flow formulizes a more robust energy function for flow field estimations by intelligently combining the ANCC cost function [25] and cost volume filter [11], which is solved very efficiently in an unified computational framework (Sec. 3.2). Unlike conventional ANCC [25], our approach first approximates it as only considering a source guidance image, which enables us to apply fast EAF very efficiently without performance loss (Sec. 3.3.1). Under constructed cost volumes, a cost volume filtering for the optimization is followed (Sec. 3.3.2), where the weights for adaptive support aggregation is also re-used as one from the cost computation. To overcome a computational bottleneck from a large label search space, we propose the PMF-like search space sampling scheme. In each iteration, estimated flow fields are considered as a new guidance image for the asymmetric ANCC cost computation and the cost filtering, which simultaneously enhances a matching quality and boosts a convergence (Sec. 3.3.2). Fig. 2 illustrates the ANCC flow framework.

3.2. Our Computational Model

Similar to cost volume filter [11] or PMF [13], we employ a local window-based cost aggregation scheme to provide a reliable matching performance with a very low computational time. In particular,
Propagate a set of labels $L_y$ randomly sampled from neighboring segments to the segment $S_m$.

Update an intermediate flow $l^*$ randomly to each superpixel $S_m$.

while not converged do

for $m = 1: M$ do

4: Propagate a set of labels $L_y$ randomly sampled from neighboring segments to the segment $S_m$.

5: Compute $G^{(c)}_{i,j}$ and $G^{(d)}_{i,j}$ for all pixel $i$.

for $l \in L_p$ do

6: Construct a cost slice $f_j$ such that $j = i - l_i$.

7: Estimate $\tilde{\Phi}^l(i, l)$ using $(4)$.

8: Estimate $\mathcal{C}(i, l)$ using $(2)$ with $l^*$.

end for

9: Update an intermediate flow $l^*$ with WTA in $(3)$.

10: Randomly sample $q \in S_m$ for defining $L_q$.

11: Update $\mathcal{C}(i, l)$ by following Step $6-9$ for all $l \in L_q$.

end for

end while

With some arithmetic derivations, $(4)$ can be decomposed as

$$
\frac{G_{i,j} - G_{i, l} \cdot G_{i,j}}{\sqrt{G_{i,j}^2 - (G_{i, l})^2}}
$$

(5)

where $G_{i,j} = \sum_{l'} \omega_{i,l'}f_{i,j}^2$, $G_{i, l} = \sum_{l'} \omega_{i,l'}f_{i,l'}f_{i,j}$, and $G_{i,j} = \sum_{l'} \omega_{i,l'}f_{i,l'}^2$. Similar to efficient computation scheme in [15], $(5)$ can be computed very efficiently using a constant-time EAF [31, 28].

Furthermore, to improve a robustness and discriminative power, our final cost function makes use of an evolving guidance aggregation in a way that the previously estimated flow field is considered as a guidance for the adaptive support aggregation in $(5)$ such that

$$
\tilde{\Phi}^l(i, l) = \frac{G_{i,j} - G_{i, l} \cdot G_{i,j}}{\sqrt{G_{i,j}^2 - (G_{i, l})^2}}
$$

(6)

where $G_{i,j}^c, G_{i,j}^d, G_{i,j}^l$, and $G_{i,j}^l$ are computed with $\omega_{i,l'}$, which is defined as edge-aware weights from previously estimated label $l$.

3.3.2. PMF based cost optimization

Our asymmetric ANCC cost in $(6)$ can be computed very efficiently compared to original ANCC in $(4)$, but its computational time still depends on search range size $|L|$. To overcome this limitation, we employ label search space sampling strategy in PMF [13]. In the optimization, for the evolving guidance aggregation, our approach utilizes the previous flow field to construct edge-aware weights. We first decompose the image $f$ as superpixel $S_f = \{S_m\} \cup S_m = \mathcal{I}$ and $\forall m \neq n, S_m \cap S_n \neq \emptyset, m = 1, ..., M$, where $M$ is the number of superpixels. A random label is initially assigned to each node, and we iterate two search strategies in an interleaved manner, i.e., neighborhood propagation and random search.

In neighborhood propagation step, for a current segment $S_m$, we denote its set of spatially adjacent neighbors as $\{S_n\}$, and candidate pixels $p \in S_n$ are then randomly sampled from every neighboring segment. A set of current best labels $L_p = \{l_p\}$ is then retrieved. For $l \in L_p$, an asymmetric ANCC cost in $(5)$ and cost volume filtering in $(2)$ are computed sequentially. After the preceding propagation step, in random search step, we randomly pick a reference pixel $q \in S_m$ to promote the label propagation within a segment. After defining a set of labels $L_q = \{l_q\}$, $(5)$ and $(2)$ are computed again for pixels $i \in S_m$. After each iteration, the intermediate flow field is inferred using $(3)$, and is applied to next iteration as an evolving guidance aggregation, which will boost matching performances on each iteration. Algorithm 1 summarizes our ANCC flow.

### 3.3.3. Effects of evolving guidance aggregation

Fig. 3 shows convergence analysis of ANCC flow. In order to analyze only effects of evolving guidance aggregations in cost computation and cost aggregation, not PMF itself [13], disparity maps from the cost computation with WTA and its corresponding cost filtering with WTA are estimated with fixed $L_p$ and $L_q$. As shown in Fig. 3, an evolving guidance aggregation dramatically improves matching performances in cost computation and cost filtering, simultaneously. It further enables boosting very fast convergence.

### 3.4. Extension to Geometric-Invariant Flow Field Estimation

By properly defining the search label space $l_i \in \mathcal{L}$, our ANCC flow can be applied to general dense correspondence problems. For stereo matching, $l_i$ is defined to assign a disparity $d$ to pixel $i$, where $v = 0$. For optical flow estimation, $l_i$ is defined to assign a 2-D vector field for $[u, v]$. More challenging, for general image matching scenarios, where there exist not only translation fields but also geometrically variations fields, e.g., scale and rotation, it is hard to directly define $l_i$ due to too many possible candidates. Instead, we employ a randomized global transform approach [32]. Specifically, based on initial sparse feature matching and RANSAC-based global transform inference [33], we estimate global transform candidates $T_c$ between multiple images, which can deal with scale and rotation field very efficiently. Using $T_c$, search label space are defined in such a way that $l_i = T_c(i)$ for all $c$, where $T_c(i)$ means that pixel $i$ is applied by global transform $T_c$. In this case, $|\mathcal{L}| = |T_c|$.}

### 3.5. Computational Complexity Analysis

Given an image size $|\mathcal{I}|$, the label space size $|\mathcal{L}|$, the number of iteration $K$, and the aggregation window size $|F|$, the computational complexity the ANCC flow framework on the brute-force implementation becomes $O(K|\mathcal{I}||\mathcal{L}|^4|F|)$. With efficient computation model with fast constant-time EAF, our approach removes the complexity dependency on the aggregation window size $|F|$, i.e., $O(K|\mathcal{I}||\mathcal{L}|)$.
4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Experimental Environments

In experiments, the ANCC flow was implemented with the following same parameter settings for all datasets. For an EAF, we employed the guided filter (GF) [28] with radius $r = 9$ and the smoothness parameter $\varepsilon = 0.009$. The number of superpixels is set to about 500 using the SLIC [35]. We implemented the ANCC flow in C++ on Intel Core i7-3770 CPU at 3.40 GHz. The computational complexity of ANCC flow compared to other methods was evaluated in Table 1.

ANCC flow was compared to state-of-the-art matching methods, e.g., cost filter [11], PatchMatch [10], and SIFT [16], DAISY [17], and DASC [15] combined with SIFT flow [36] optimization. Furthermore, we compared our approach with NCC [6] and ANCC [34] with GC [9] optimization. For geometric-invariant flow estimations, we evaluated SID [37], SegSID [37], and SSF [38].

4.2. Middlebury Stereo Benchmark

We first evaluate our ANCC flow framework in Middlebury stereo benchmark containing illumination and exposure variations [6]. In the experiments, the illumination (or exposure) combination ‘1/3’ indicates that two images were captured under the 1st and 3rd illumination (exposure) conditions. For a quantitative evaluation, we measured the bad-pixel error rates in non-occluded areas [6].

Fig. 4 shows depth maps for severe illumination and exposure variations, and Fig. 5 shows average bad matching error rates. As expected, without robust cost functions, the cost filter [11] and PatchMatch [10] cannot provide reliable correspondence performances. Matching performances of SIFT flow [36] combined with robust cost functions [17, 18, 20, 15] are limited on edge-discontinuity regions since they provide limited discriminative power. Furthermore, their computational time was very high. Unlike these conventional methods, our ANCC flow achieved the best results both quantitatively and qualitatively. By using an leveraging guidance aggregation, the matching performance of ANCC flow was highly enhanced.

4.3. DIML Benchmark

We then evaluate our ANCC flow framework in recently published DIML benchmark [15], captured as 10 geometry image sets by combining geometric variations of viewpoint, scale, and rotation, and each image set consists of images taken under 5 different photometric variation pairs including illumination, exposure, flash-noflash, blur and noise. To evaluate the performance quantitatively, we computed the label transfer accuracy (LTA) [39, 15].

Fig. 6 shows qualitative evaluation results, and Table 2 shows average LTA error rates on DIML benchmark [15]. (a),(b) image pairs, warped color images from correspondences of (c) SIFT flow [36], (d) SID [37], and (e) ANCC flow.
6. REFERENCES


[8] D. Sun, S. Roth, and M. Black, “Secret of optical flow estima-


[23] S. Kim, S. Ryu, B. Ham, B. Kim, and K. Sohn, “Local self-


[27] X. Shen, L. Xu, Q. Zhang, and J. Jia, “Multi-modal and multi-


[33] Q. Tran, T. Chin, G. Carneiro, M. S. Brown, and D. Suter, “In defence of ransac for outlier rejection in deformable registra-

[34] Y. Heo, K. Lee, and S. Lee, “Joint depth map and color consist-


[37] I. Kokkinos and A. Yuille, “Scale invariance without scale se-
