# Extending Layered Models to 3D Motion Supplementary Material

Dong Lao and Ganesh Sundaramoorthi

KAUST
{dong.lao,ganesh.sundaramoorthi}@kaust.edu.sa

## 1 Derivation of Equation (3) to (6) in the Paper

In this section, Eq(x) refers to equation (x) in the paper, while (x) refers to equation (x) in the supplementary material.

### 1.1 Eq(3)

We optimize in $f_i$ given estimates of the other variables. $f_i$ appears only in the first term of $E_{app}$. Following the notations in the paper,

$$E = \sum_{t,i} \int_{\tilde{R}_{it}} |I_t(x) - f_i(w_{it}^{-1}(x))|^2 \,\mathrm{d}x \tag{1}$$

$$= \sum_{i} \int_{R_i} \sum_{t} |I_t(w_{it}(x)) - f_i(x)|^2 V_{it}(w_i(x)) J_{it}(x) \,\mathrm{d}x \tag{2}$$

Let $L = \sum_t [|I_t(w_{it}(x)) - f_i(x)|^2 V_{it}(w_i(x)) J_{it}(x)]$. By Euler-Lagrange equation, the optimizer satisfies $\frac{\partial L}{\partial f_i} = 0$. Thus

$$2\sum_{t} [(I_t(w_{it}(x)) - f_i(x)) V_{it}(w_{it}(x)) J_{it}(x)] = 0 \tag{3}$$

$$f_i(x) = \frac{\sum_t I_t(w_{it}(x)) V_{it}(w_{it}(x)) J_{it}(x)}{\sum_t V_{it}(w_{it}(x)) J_{it}(x)}, \quad x \in R_i, \tag{4}$$

### 1.2 Eq(4) and Eq(5)

Similar to (1), by a change of variables the energy can be defined in each $R_i$. Following the notations in the paper,

$$E_{app} = \sum_{t,i} \int_{\tilde{R}_{it}} |I_t(x) - f_i(w_{it}^{-1}(x))|^2 \,\mathrm{d}x - \int_{\tilde{R}_{it}} \beta_t(x) \log p_i(I_t(x)) \,\mathrm{d}x \tag{5}$$

$$= \sum_{i} \int_{R_i} \sum_{t} [|\tilde{I}_t(x) - \tilde{f}_i(x)|^2 - \beta_t(x) \log p_i(\tilde{I}_t(x))] J_{it}(x) \tilde{V}_i(x) \,\mathrm{d}x \tag{6}$$

From this expression, the problem can be formulated as a region competition problem, which was described in [1]. Following $\nabla_{\partial R} \int_R f(x)dx = fN$ , where N is the outward normal to $\partial R$ [1], it is straightforward that the gradient flow at $\partial R_i$ is

$$\begin{aligned}&\nabla_{\partial R_i} E_{app}\\&= \sum_t \left\{ [|\tilde{I}_{it} - f_i|^2 - \beta_t \log p_i(\tilde{I}_{it})] - [|\tilde{I}_{jt} - \tilde{f}_j|^2 - \beta_t \log p_j(\tilde{I}_{jt})] \right\} J_{it} \tilde{V}_i N_i \end{aligned} \quad (7)$$

Similarly we can compute the gradient flow of $E_{reg}$. Finally by $E = E_{app} + E_{reg}$ we have Eq(4)

$$\nabla_{\partial R_i} E = \sum_t \left[ |\tilde{I}_{it} - f_i|^2 - |\tilde{I}_{jt} - \tilde{f}_j|^2 - \beta_t \log \frac{p_i(\tilde{I}_{it})}{p_j(\tilde{I}_{jt})} + \alpha \kappa_i \right] J_{it} \tilde{V}_i N_i + \gamma N_i. \quad (8)$$

The compuation of Eq(5) is the same.

### 1.3 Eq(6)

The derivation of Eq(6) follows the results from [2]. The Sobolev gradient $G_{it} = \nabla_{w_{it}} E$ with respect to $w_{it}$ is a linear combination of a translation component $\operatorname{avg}(G_{it})$ and a deformation component $\tilde{G}_{it}(x)$, where $G_{it} = \operatorname{avg}(G_{it}) + \alpha \tilde{G}_{it}$ and $\operatorname{avg}(\tilde{G}_{it}) = 0$. Consider the only term in the energy function Eq(1) that contains $w$:

$$E = \int_{\tilde{R}_{it}} |I_t(x) - f_i(w_{it}^{-1}(x))|^2 \, \mathrm{d}x. \quad (9)$$

Note that $\tilde{R}_{it} = w_{it}(R) \cap \{\tilde{V}_i = 1\}$, where $\tilde{V}_i = V_{it} \circ w_{it}$ denotes the visibility. For simplicity of the notation, in this section we omit subscripts and write (9) as:

$$E = \int_{w(R)} |I(x) - \hat{f}(x)|^2 \tilde{V}(x) \, \mathrm{d}x. \quad (10)$$

Computing the variation of $E$ with respect to a perturbation $h$ of $w$ yields

$$\begin{aligned}\mathrm{d}E(w) \cdot h = &\int_{\partial w(R)} [I(x) - \hat{f}(x)]^2 \cdot \tilde{N} \tilde{V} h(x) \, \mathrm{d}s(x) \\ &+ \int_{w(R)} \nabla \hat{f}(x) \cdot [I(x) - \hat{f}(x)]^T \tilde{V} h(x) \, \mathrm{d}x. \end{aligned} \quad (11)$$

By defination of functional derivative, $\mathrm{d}E(w) \cdot h = \langle G, h \rangle_w$, where $G = \nabla_w E$ is the gradient with respect to the Sobolev inner product $\langle a, b \rangle = \operatorname{avg}(a) \cdot \operatorname{avg}(b) + \alpha \int_R \nabla a(x) \cdot \nabla b(x) \, \mathrm{d}x$ [2]. For simplicity we eliminate $\alpha$. By integrating by parts,

$$
\begin{aligned}
\mathrm{d}E(w)\cdot h &= \operatorname{avg}(G)\cdot\operatorname{avg}(h) + \int_{w(R)} \nabla\tilde{G}\cdot\nabla h\,\mathrm{d}x \\
&= \operatorname{avg}(G)\cdot\operatorname{avg}(h) + \int_{\partial w(R)} \nabla\tilde{G}(x)\cdot\tilde{N}h(x)\,\mathrm{d}s(x) \\
&\quad - \int_{w(R)} \Delta\tilde{G}(x)h(x)\,\mathrm{d}x.
\end{aligned} \tag{12}
$$

By comparing (11) and (12), we have

$$
\begin{cases}
-\Delta\tilde{G}(x) = F(x) & x\in w(R) \\
\nabla\tilde{G}(x)\cdot\tilde{N} = |I-\hat{f}|^2\tilde{V}\tilde{N} & x\in\partial w(R) \\
\operatorname{avg}(G) = \operatorname{avg}(F)
\end{cases}, \quad F = \nabla\hat{f}[I-\hat{f}]^T\tilde{V}. \tag{13}
$$

# 2 Full Details of Numerical Implementation

## 2.1 Numerical Implementation of Algorithm

We provide more details of our numerical scheme for the evolution of the flattened regions $R_i$ and the segmentations $\tilde{R}_{it}$. We implement the evolutions of their boundaries using a standard narrowbanding level set method [3]. Each of the regions $R_i$ are represented by a level set function $\phi_i^\tau$ and the regions are related to the level set as $R_i^\tau = \{\phi_i^\tau = 1/2\}$ where $\tau$ is the time parameter of the evolution. Similarly, $\tilde{R}_{it}^\tau$ are represented with level sets $V_{it}$. The corresponding level set evolutions are shown in Algorithm 1. Step sizes are chosen to satisfy the CFL conditions.

# 3 Extended DAVIS Results

## 3.1 Video Sample Results

See the provided movie files for representative results on the DAVIS dataset.

## 3.2 Full Results on Davis 2016 dataset

Tab. 1 presents the per-sequence F-measure on Davis dataset. In 15 / 50 sequences our methods achieves the best performance.

## 3.3 Extended Discussion of Success / Failures

Sequences labeled green are the ones in which the objects are occluded and split into more than one parts. Our result shows strong performance in these sequences. Occlusion does not affect the segmentation accuracy since no depth

| Sequence | Ours | ARP[4] | LVO[5] | FSEG[6] | LMP[7] | SFL[8] | FST[9] | CUT[10] | NLC[11] |
|---|---|---|---|---|---|---|---|---|---|
| Bear | 0.843 | **0.894** | - | 0.869 | 0.665 | - | 0.860 | - | 0.850 |
| Blackswan | 0.513 | 0.912 | 0.774 | 0.767 | 0.594 | **0.939** | 0.736 | 566 | 0.820 |
| BMX-Bumps | **0.791** | 0.574 | - | 0.451 | 0.522 | - | 0.348 | - | 0.734 |
| BMX-Trees | **0.913** | 0.678 | 0.674 | 0.599 | 0.657 | 0.605 | 0.348 | 0.505 | 0.330 |
| Boat | 0.550 | 0.408 | - | **0.643** | 0.480 | - | 0.197 | - | 0.036 |
| Breakdance | 0.513 | **0.741** | 0.340 | 0.419 | 0.503 | 0.204 | 0.411 | 0.340 | 0.661 |
| Breakdance-Flare | 0.783 | **0.859** | - | 0.794 | 0.850 | - | 0.694 | - | 0.808 |
| Bus | **0.746** | 0.586 | - | 0.563 | 0.547 | - | 0.584 | - | 0.406 |
| Camel | 0.882 | 0.902 | 0.889 | 0.835 | 0.803 | 0.812 | 0.590 | **0.922** | 0.719 |
| Car-roundabout | 0.643 | 0.621 | 0.745 | **0.894** | 0.637 | 0.892 | 0.625 | 0.547 | 0.250 |
| Car-Shadow | 0.715 | 0.692 | **0.945** | 0.940 | 0.771 | 0.937 | 0.540 | 0.776 | 0.546 |
| Car-Turn | 0.690 | 0.763 | - | **0.912** | 0.676 | - | 0.731 | - | 0.634 |
| Cows | 0.859 | 0.860 | **0.885** | 0.816 | 0.773 | 0.693 | 0.681 | 0.785 | 0.807 |
| Dance-Jump | **0.618** | 0.602 | - | 0.463 | 0.459 | - | 0.462 | - | 0.567 |
| Dance-Twirl | **0.906** | 0.797 | 0.789 | 0.651 | 0.594 | 0.583 | 0.471 | 0.715 | 0.365 |
| Dog | 0.729 | 0.713 | 0.837 | **0.864** | 0.787 | 0.941 | 0.659 | 0.678 | 0.707 |
| Dog-Agility | 0.280 | 0.266 | - | **0.569** | 0.337 | - | 0.265 | - | 0.551 |
| Drift-Chicane | 0.239 | **0.889** | 0.711 | 0.654 | 0.771 | 0.146 | 0.731 | 0.710 | 0.312 |
| Drift-Straight | 0.441 | 0.539 | 0.721 | 0.652 | 0.534 | **0.827** | 0.470 | 0.551 | 0.385 |
| Drift-Turn | 0.283 | 0.645 | - | **0.744** | 0.464 | - | 0.442 | - | 0.185 |
| Elephant | 0.598 | **0.660** | - | 0.659 | 0.627 | - | 0.569 | - | 0.251 |
| Flamingo | 0.800 | **0.838** | - | 0.812 | 0.778 | - | 0.763 | - | 0.610 |
| Goat | 0.447 | 0.746 | 0.766 | 0.799 | 0.707 | **0.806** | 0.400 | 0.479 | 0.133 |
| Hike | 0.894 | **0.944** | - | 0.759 | 0.891 | - | 0.918 | - | 0.943 |
| Hockey | 0.877 | 0.767 | - | 0.676 | **0.878** | - | 0.584 | - | 0.808 |
| Horsejump-High | 0.880 | 0.882 | **0.888** | 0.658 | 0.882 | 0.748 | 0.621 | 0.690 | 0.881 |
| Horsejump-Low | **0.832** | 0.781 | - | 0.717 | 0.794 | - | 0.490 | - | 0.659 |
| Kite-Surf | 0.295 | 0.377 | **0.523** | 0.322 | 0.473 | 0.397 | 0.346 | 0.272 | 0.448 |
| Kite-Walk | **0.679** | 0.421 | - | 0.340 | 0.592 | - | 0.561 | - | 0.662 |
| Libby | **0.843** | 0.735 | 0.819 | 0.674 | 0.796 | 0.824 | 0.718 | 0.359 | 0.748 |
| Lucia | **0.935** | 0.855 | - | 0.784 | 0.883 | - | 0.568 | - | 0.872 |
| Mallard-Fly | 0.323 | 0.614 | - | **0.711** | 0.649 | - | 0.633 | - | 0.661 |
| Mallard-Water | 0.111 | 0.491 | - | **0.756** | 0.214 | - | 0.079 | - | 0.692 |
| Motocross-Bumps | **0.819** | 0.728 | - | 0.661 | 0.699 | - | 0.610 | - | 0.560 |
| Motocross-Jump | 0.513 | **0.646** | 0.630 | 0.504 | 0.582 | 0.608 | 0.453 | 0.461 | 0.303 |
| Motorbike | 0.743 | 0.628 | - | 0.418 | **0.782** | - | 0.584 | - | 0.571 |
| Paragliding | **0.949** | 0.745 | - | 0.289 | 0.905 | - | 0.675 | - | 0.744 |
| Paragliding-Launch | **0.254** | 0.193 | 0.221 | 0.176 | 0.253 | 0.187 | 0.185 | 0.201 | 0.243 |
| Parkour | 0.485 | 0.843 | 0.871 | 0.779 | 0.789 | 0.846 | 0.478 | 0.442 | **0.916** |
| Rhino | **0.875** | 0.823 | - | 0.760 | 0.684 | - | 0.634 | - | 0.431 |
| Rollerblade | **0.945** | 0.904 | - | 0.694 | 0.761 | - | 0.411 | - | 0.868 |
| Scotter-Black | 0.433 | 0.587 | 0.575 | 0.534 | 0.563 | **0.615** | 0.395 | 0.434 | 0.228 |
| Scotter-Gray | 0.437 | 0.567 | - | 0.531 | **0.609** | - | 0.321 | - | 0.466 |
| Soapbox | 0.768 | 0.766 | **0.821** | 0.520 | 0.709 | 0.721 | 0.355 | 0.597 | 0.658 |
| Soccerball | 0.876 | 0.855 | - | 0.867 | 0.851 | - | **0.900** | - | 0.855 |
| Stroller | 0.758 | **0.878** | - | 0.663 | 0.561 | - | 0.558 | - | 0.874 |
| Surf | 0.741 | **0.906** | - | 0.823 | 0.434 | - | 0.445 | - | 0.673 |
| Swing | **0.830** | 0.699 | - | 0.628 | 0.756 | - | 0.491 | - | 0.778 |
| Tennis | 0.880 | 0.843 | - | 0.764 | 0.838 | - | 0.567 | - | **0.927** |
| Train | 0.765 | **0.879** | - | 0.570 | 0.777 | - | 0.660 | - | 0.521 |

**Table 1. Per-Sequence F-measure on Davis.**

**Algorithm 1** *Layered optimization numerical implementation*

1: Input: Initialization for the flattened representations $R_i, f_i$
2: **repeat** // *update the flattened representations, warps and segmentations*
3: For all $i$ and $t$, update $w_{it}$ performing gradient descent Eq(6) until convergence
4: For all $i$, compute $f_i$ by Eq(3)
5: For all $i$, update $R_i$ by one step in negative gradient direction Eq(4):

$$\phi_i^{\tau+\Delta\tau}(x) = \phi_i^{\tau}(x) - \Delta\tau \left[ \sum_t [|\tilde{I}_{it} - f_i|^2 - |\tilde{I}_{jt} - \tilde{f}_j|^2] J_{it} \tilde{V}_i + \gamma \right] |\nabla \phi_i^{\tau}(x)| + \alpha |\nabla \phi_i^{\tau}(x)| \sum_t \mathrm{div} \left[ \frac{\nabla V_{it}^{\tau}(x)}{|\nabla V_{it}^{\tau}(x)|} \right] J_{it}$$

for all $x$ in a narrowband of $\{\phi_i^{\tau} = 1/2\}$
6: For all $t$, update the $V_{it}$ by one step in negative gradient direction Eq(5):

$$V_{it}^{\tau+\Delta\tau}(x) = V_{it}^{\tau}(x) - \Delta\tau \sum_t \left[ |I_t - \hat{f}_i|^2 - |I_t - \hat{f}_j|^2 - \beta_t \log \frac{p_i(I_t)}{p_j(I_t)} \right] |\nabla V_{it}^{\tau}(x)| + \alpha |\nabla V_{it}^{\tau}(x)| \mathrm{div} \left[ \frac{\nabla V_{it}^{\tau}(x)}{|\nabla V_{it}^{\tau}(x)|} \right]$$

for all $x$ in a narrowband of $\{V_{it}^{\tau} = 1/2\}$
7: $\tilde{R}_{it} = \{V_{it}^{\tau} = 1/2\}$, $R_i = \{\phi_i^{\tau} = 1/2\}$
8: **until** the energy $E$ converges

ordering is required, an advantage of our method. See Fig. 1 for examples of the segmentation results.

Sequences labeled red are the ones containing strong irregular motion in the background, which may be caused by dynamic background (e.g water waves) or unlabeled moving object. Motion segmentation schemes correctly detect this as motion, but the dataset doesn't consider these as objects. See Fig. 2 for example failure cases.

## 4 Extended MIT Results

**Visualized Result on MIT Layer Dataset** Fig. 3 presents the visualized result on MIT Layer Dataset. In most cases layers are correctly recovered including the ones containing 3D non-planar motion and self-occlusion.

**Fig. 1. Successful cases.**

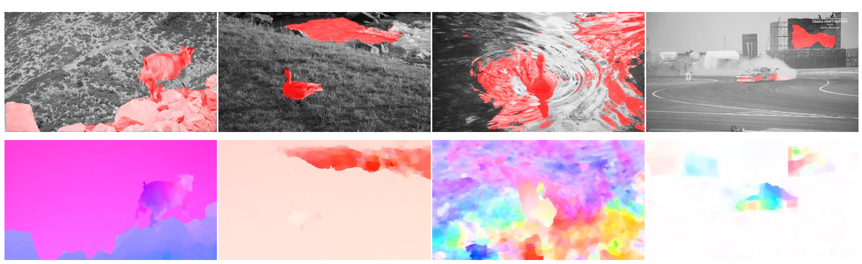

**Fig. 2. Failure cases.** *[Up]: Faulty segmentation results. [Down]: Frame by frame optical flow. Strong and Irregular motion in the background*


**Fig. 3. Results on MIT Layer dataset.** *For each sequence, the results on three different frames are presented.*

## References


1. Zhu, S.C., Yuille, A.: Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation. IEEE transactions on pattern analysis and machine intelligence **18**(9) (1996) 884–900
2. Yang, Y., Sundaramoorthi, G.: Shape tracking with occlusions via coarse-to-fine region-based sobolev descent. IEEE transactions on pattern analysis and machine intelligence **37**(5) (2015) 1053–1066
3. Sethian, J.A.: Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science. Volume 3. Cambridge university press (1999)
4. Jun Koh, Y., Kim, C.S.: Primary object segmentation in videos based on region augmentation and reduction. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (July 2017)
5. Tokmakov, P., Alahari, K., Schmid, C.: Learning video object segmentation with visual memory. arXiv preprint arXiv:1704.05737 (2017)
6. Jain, S.D., Xiong, B., Grauman, K.: Fusionseg: Learning to combine motion and appearance for fully automatic segmention of generic objects in videos. arXiv preprint arXiv:1701.05384 **2**(3) (2017) 6
7. Tokmakov, P., Alahari, K., Schmid, C.: Learning motion patterns in videos. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2017) 531–539
8. Cheng, J., Tsai, Y.H., Wang, S., Yang, M.H.: Segflow: Joint learning for video object segmentation and optical flow. In: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE (2017) 686–695
9. Papazoglou, A., Ferrari, V.: Fast object segmentation in unconstrained video. In: Computer Vision (ICCV), 2013 IEEE International Conference on, IEEE (2013) 1777–1784
10. Keuper, M., Andres, B., Brox, T.: Motion trajectory segmentation via minimum cost multicuts. In: Computer Vision (ICCV), 2015 IEEE International Conference on, IEEE (2015) 3271–3279
11. Faktor, A., Irani, M.: Video segmentation by non-local consensus voting. In: BMVC. Volume 2. (2014) 8