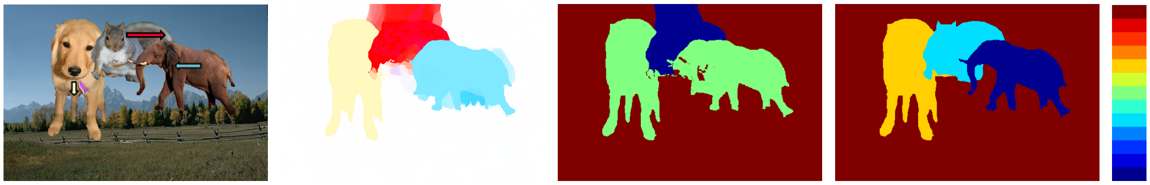

Figure 2. A failure case for the **Layers++** method [27]. Left to right: first image in a pair (arrows show motion direction and their length indicates motion magnitude); initial flow estimate, color coded as in [1]; segmentation by **Layers++**; segmentation with our proposed **nLayers** method, which automatically determines the number of layers, their depth ordering, and is able to make large changes to the initial flow field to reach a good solution. On the far right is a color key for the ordering of depth layers (blue is close and red is far).

However, our gradient-based inference algorithm is susceptible to local optima, resulting in errors in the estimated scene structure and flow field, as illustrated in Figure 2.

Overcoming such limitations requires an optimization method that can make large changes to the solution at a single step, a task more suitable for discrete optimization. Hence we propose a discrete layered model based on a sequence of ordered Markov random fields (MRFs). This model, unlike standard Ising/Potts MRFs, cannot be directly solved by "off-the-shelf" optimizers, such as graph cuts. Therefore we develop a sequence of non-standard moves that can simultaneously change the states of several binary MRFs. We also embed continuous flow estimation into the discrete framework to adapt the state space to estimate sub-pixel motion. The resultant discrete-continuous scheme enables us to infer the number of layers and their depth ordering automatically for a sequence.

We evaluate our layer segmentation using the MIT human-assisted motion annotation dataset [18]. Our method produces semantically more meaningful segmentations that are also quantitatively more consistent with human labeled ground truth than the continuous-only **Layers++** method. With a reliable layer segmentation and the relative depth ordering obtained with the discrete method, we initialize the more precise **Layers++** continuous model of optical flow. The discrete-continuous approach gives a concrete improvement over a purely continuous optimization that can easily become trapped in local optima.

Like many approaches, our previous work [27] considers optical flow between only two frames. Unfortunately, with only two frames, depth ordering at occlusion boundaries is fundamentally ambiguous [7]. Critically, our approach is formulated to estimate optical flow over time. By estimating layer segmentations over three or more frames we obtain a reliable depth ordering of the layers and more accurate motion estimates. At the time of writing, the proposed method is ranked first in AAE and fourth in EPE on the Middlebury optical flow benchmark.

In summary, our contributions include *a)* formulating a discrete layered model based on a sequence of ordered Ising MRFs and devising a set of non-standard moves to optimize it; *b)* formulating methods for automatically determining the number of layers and their depth ordering for a given sequence; *c)* concretely improving layer segmentation on a set of real-world sequences; *d)* demonstrating the benefits of using more frames for optical flow estimation on the Middlebury optical flow benchmark.

2. Previous Work

Layered optical flow. Most layered approaches assume a parametric motion for each layer [11, 13, 16, 31] which is too restrictive to capture the motion of natural scenes. Weiss [32] addresses this by allowing smooth motions in the layers. In [27] we extend this to impose global coherence via an affine motion field while modeling local non-smooth deformation from affine with a robust MRF. Both methods adopt continuous optimization methods that do not reason about the number or ordering of layers.

Like us, Jepson *et al.* [12] decompose a scene into overlapping layers, reasoning about the number of layers, and determining the depth order. While their method models layer support with parametric regions we allow much more varied layers. Weiss and Adelson [33] incorporate static image cues into the layered segmentation and estimate the number of layers under fairly weak assumptions. Torr *et al.* [28] use a Bayesian decision making framework to determine the number of approximately planar layers but do not infer the depth ordering. Our depth ordering formulation is similar to flexible sprites [13] and their extensions [16], but we use more flexible motion models.

Occlusion. Reasoning about occlusion in image sequences dates to the mid 1970's and early 1980's; a full review is beyond our scope. Early authors (*e.g.* [21]) note that occlusion boundaries move with the occluding surface but the first explicit statement that this requires three frames to compute seems to be by Darrell and Fleet [7]. We illustrate this in Figure 3 because we can not find a clear description in the literature.

This simple fact is a key reason why two-frame optical flow estimation is fundamentally limited. In a layered model, inferring the wrong depth order results in significant errors at motion boundaries. The idea of using three or more frames has been embodied in recent methods for computing motion boundaries and depth order [3, 8] but appears missing from recent dense flow estimation methods.

Estimating flow over time. Again, estimation of flow