Implicit Representations for Recovery and Reconstruction of Motion Layers

[+]Rong Zhang [&]Nikos Paragios [+]Dimitris Metaxas

[+] Department of Computer Science
Rutgers University
110 Frelinghuysen Road
Piscataway, NJ 08854, USA

[&] Real Time Vision & Modeling
Siemens Corporate Research
755 College Road East
Princeton, NJ 08540, USA

Abstract

In this paper, we propose a level set approach to decompose the image plane into segments with consistent apparent motion and recover parametric models that best describe such a motion. The number of motion planes and initial motion parameters are obtained by robust block-wise optical flow estimation, followed by non-parametric clustering in the motion parameter space. Implicit representations are used to further improve the estimation and to segment the image plane according to consistent motion. Our approach integrates a smoothness constraint, a spatial-temporal motion constraint and a color-driven segmentation module within the multiphase level set framework. Therefore, our method is able to recover the different topology of the motion layers and provide smooth motion segmentation that is consistent with the geometry of the scene. Experimental results on both natural and synthetic image sequences demonstrate the potential of the proposed technique.

1. Introduction

The segmentation of an image sequence into regions with homogeneous motion is a challenging task in video processing [2, 4, 22]. The motion segmentation results can be used for various purposes such as video-based surveillance and action recognition. In addition, it can be considered for video compression [2] since the motion model and the corresponding supporting layers provide a compact representation of the scene.

Motion/displacement is a well-defined measurement in the real world. On the other hand, one can claim that recovering the corresponding quantity in the image plane is a difficult task. Optical flow calculation [9] is equivalent with the estimation of a motion displacement vector for each pixel of the image plane that satisfies the visual constancy constraint. Such a task refers to an ill-posed problem where the number of unknown variables exceeds the number of constraints. The use of smoothness constraints [20] and other sophisticated techniques were considered to address such an issue.

Parametric motion models are an alternative to dense optical flow estimation [14]. The basic assumption of such a technique is that for an image block, the 2D motion in the image plane can be modeled using a parametric transformation. Such assumption is valid when the block refers to a projection of 3D patch with a constant depth from the camera position.

The objective of the this work is to recover different planar surfaces, or motion layers, and the motion parameters describing their apparent displacements. In the literature, a K-mean clustering algorithm [22] on the motion estimates, or a minimum description length (MDL) [2] were considered to determine the number of motion planes. In the later case, the extraction is done according to a maximum likelihood criterion, followed by optimization by the Expectation-Maximization algorithm [8, 2].

In this paper we consider a level set formulation [15] for motion estimation grouping into multiple layers of scenes observed by moving camera. To this end, first a block-based robust parametric motion estimation is considered. Such estimates define a multi-cluster probability density function in the space of motion parameters. The number of motion layers is obtained using a non-parametric clustering technique [5]. Then, a variational level set formulation [16] is considered [21] to recover the motion layers and the optimal estimates of their parameters. The objective function consists of a smoothness term, an optical flow component and a visual grouping term [17]. The use of calculus of variations with a gradient descent method is used to recover the lowest potential of the cost function.

The remainder of this paper is organized as follows. In section 2 we present details on the estimation of the number of layers. Decomposition of the image into motion layers and estimation of the corresponding motion parameters are presented in section 3. Discussion and comparison with existing techniques are presented in section 4.

2. Initial Conditions

Rigid, similarity, affine, projective and quadratic motion models were considered for parametric motion estimation. Among these methods, affine motion models [14] are quite popular in motion analysis since they refer to a compromise between low complexity and fair performance. Such models refer to six degrees of freedom:

$$\mathcal{A}(x,y) = \left[\begin{array}{c} \mathcal{A}^x(x,y) \\ \mathcal{A}^y(x,y) \end{array}\right] = \left[\begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array}\right] \left[\begin{array}{c} x \\ y \end{array}\right] + \left[\begin{array}{c} a_{13} \\ a_{23} \end{array}\right]. \quad (1)$$

Within our approach, we assume that the motion of each layer can be described using such a model. Under this an assumption, motion reconstruction is equivalent to recovering the number of layers, the corresponding motion parameters, and the spatial support for each layer. The number of motion planes is a critical parameter since it constrains the solution space significantly.

One can consider optimizing an objective function that consists of all terms. Such a selection might refer to a considerable complexity and uncertain stability with respect to the initial conditions. In order to cope with such constraint, and to reduce the complexity of the problem, we determined the number of motion layers and the motion planes along with the optimal motion parameters in two steps.

First, robust estimation of affine motion models over image blocks is performed. Towards determining the number of motion planes, non-parametric clustering in the space of the affine model parameters is considered. The centers of the motion clusters are also used to provide an initial estimation on the motion parameters of each layer.

2.1. Block-wise Motion Parameter Estimation

Visual consistency [9] is a constraint for the recovery of the apparent motion, or optical flow. Let us consider an image sequence $I(\mathbf{s};t)$ [$\mathbf{s} = (x,y)$] and in particular two consecutive frames, $I(\mathbf{s};\tau)$ and $I(\mathbf{s};\tau+1)$. Assume pixel $\mathbf{s}$ at time τ move to position $\mathbf{s}+(u,v)$ at time $\tau+1$, visual consistency is to assume its color does not change during this movement[1], i.e.,

$$I(\mathbf{s};\tau) = I(\mathbf{s}+(u,v);\tau+1). \quad (2)$$

One can use a first-order Taylor expansion to obtain the following term known as the optical flow constraint (OFC)

$$I_t(\mathbf{s}) + uI_x(\mathbf{s}) + vI_y(\mathbf{s}) = 0, \quad (3)$$

where time has been omitted, and I_t, I_x, I_y are the time, $x-$, $y-$ derivatives, respectively. When a parametric motion approach is considered, the above OFC can be rewritten as

$$I_t(\mathbf{s}) + \mathcal{A}^x(s)I_x(\mathbf{s}) + \mathcal{A}^y(s)I_y(\mathbf{s}) = 0. \quad (4)$$

[1]Under certain assumption for the reflections properties of the scene.

Such condition is not sufficient to recover the motion parameters. Assuming same parametric motion models are valid over a large image window, one can turn this problem into a over constraint one. Such models are valid for image blocks that correspond to 3D surfaces at a constant depth from the camera position. Particular handling is required for discontinuous image locations due to violation of the considered hypothesis.

The assumption of the parametric motion model $\mathcal{A}$ within a patch $\mathcal{W}$ can lead to the following condition:

$$(u,v) = \mathcal{A}(\mathbf{s}); \quad \mathbf{s} \in \mathcal{W}. \quad (5)$$

Obviously, recovering the motion model $\mathcal{A}$ is not trivial due to the absence of a unique solution. Therefore, we consider a variational framework where the above conditions are translated to an objective function to be minimized. The least square estimator (LSE) is a well known, simple but efficient technique to define such a function:

$$E(\mathcal{A}) = \iint_{\mathcal{W}} (I_t(\mathbf{s}) + \mathcal{A}^x(\mathbf{s})I_x(\mathbf{s}) + \mathcal{A}^y(\mathbf{s})I_y(\mathbf{s}))^2 d\mathbf{s} \quad (6)$$

The best motion model $\mathcal{A}$ is the one minimizing $E(\mathcal{A})$, which can be obtained using a gradient descent method. However, LSE are sensitive to the presence of outliers or discontinuities. Robust estimators [10] are an alternative technique to recover $\mathcal{A}$ while accounting for the presence of outliers. To this end, the error-two norm in LSE is replaced with the error metric ρ that has bounded behavior, continuous first order derivatives and certain other characteristics [13]:

$$E(\mathcal{A}) = \iint_{\mathcal{W}} \rho\left(I_t + \mathcal{A}^x I_x + \mathcal{A}^y I_y\right) d\mathcal{W}, \quad (7)$$

where $\mathbf{s}$ has been omitted to simplify the notation.

Particular attention has to be paid when selecting the form of the ρ function. Inspired by [2], we consider the following error function:

$$\rho(r) = \frac{1}{\sqrt{|\sigma|}}\left(\frac{r^2}{r^2+\sigma^2} - 1\right), \quad (8)$$

where

$$r = I_t + \mathcal{A}^x I_x + \mathcal{A}^y I_y, \quad (9)$$

is the residual error, and a robust estimate (see [2] for details) of σ is given by

$$\sigma = 1.4826 \, \mathrm{median}\left\{|r_j|\right\}. \quad (10)$$

The error function is a modified Geman-McClure function as in [2]. Such modification does not affect the affine motion parameter estimation. However, small σ will result in small values when the residue is close to zero as shown in

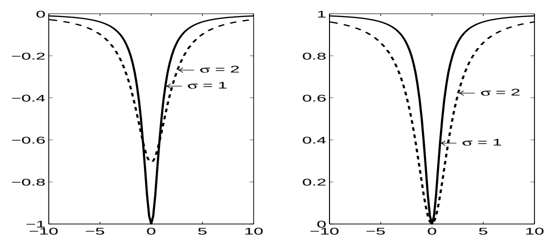


Figure 1: ρ function with different σ. Left: modified Geman-McClure function. Right: Geman-McClure function

[Figure (1)], which is necessary in the region competition part.

One can consider a variational formulation of such cost function, seeking to estimate the improvement $\Delta\mathcal{A}$ such that

$$E(\Delta\mathcal{A}) = \iint_{\mathcal{W}} \rho\left(I_t(\mathcal{A}+\Delta\mathcal{A}) + [\mathcal{A}+\Delta\mathcal{A}]^x I_x + [\mathcal{A}+\Delta\mathcal{A}]^y I_y\right) d\mathcal{W} \tag{11}$$

reaches its lowest potential. The calculus of variations with respect to $\Delta\mathcal{A}$ can lead to a flow according to the gradient descent method.

We adopt a sliding window technique to obtain a set of motion parameter estimations over the image patches [2]. Such estimates define a six-dimensional space consisting of several clusters. Under the assumption of consistent motion parameters within each motion layers, each cluster corresponds to one layer. A small number of outliers will be present due to discontinuities or image patches consisting multiple motion hypotheses.

2.2. Motion Models and Clustering

Clustering in high-dimensional spaces is a well studied problem in statistics. Parametric and non-parametric techniques can be used to solve the problem efficiently. Quite often in parametric techniques, the assumptions made on the form of distribution of the different clusters are unrealistic. This can be dealt with using non-parametric techniques such as k-means algorithm or the mean shift [5]. The latter is better suited, within our application, to automatically determine the number of clusters.

One can consider the estimates of affine motion parameters to be random samples drawn independently from a distribution with density $f(\mathcal{A})$. Such distribution is defined

[2] The parametric motion estimation of the optical flow is not important for our algorithm to this point. It is mainly used to recover the number of motion layers. Motion estimates are refined within the motion reconstruction process.

(a) (b) (c)

Figure 2: (a,b) Input Sequence, (c) Clustering on the space of motion parameters leads to four segments (the navy color region refers to discontinuities).

in the 6-D affine parameter space, which could be approximated using a kernel-based multi-variant density estimation according to a kernel K and a bandwidth h:

$$\hat{f}(\mathcal{A}) = \frac{1}{n\,h^d} \sum_{i=1}^{n} K\left(\frac{\mathcal{A}-\mathcal{A}_j}{h}\right), \tag{12}$$

where d is the dimension of the affine transformation vector and n is the number of image blocks on which motion models were estimated. The use of the Epanechnikov kernel [5] K_E is an optimal selection according to the mean integrated square error criterion:

$$K_E(\mathcal{A}) = \begin{cases} \frac{1}{2}c_d^{-1}(d+2)(1-||\mathcal{A}||^2) & ||\mathcal{A}|| \leq 1 \\ 0 & otherwise \end{cases}, \tag{13}$$

where c_d refers to the volume of the unit d-dimensional sphere. The estimation of the density gradient is equivalent to estimating the gradient of the kernel, which leads to the following condition:

$$\nabla\hat{f}(\mathcal{A}) = \frac{n_{\mathcal{A}}}{n(h^d c_d)} \frac{d+2}{h^2} \left(\frac{1}{n_{\mathcal{A}}} \sum_{\mathcal{A}_j \in S_h(\mathcal{A})} [\mathcal{A}_j - \mathcal{A}] \right) \tag{14}$$

defined in a hyper-sphere $S_h(\mathcal{A})$ centering at $\mathcal{A}$ with a volume of $h^d c_d$, where $n_{\mathcal{A}}$ is the number of samples within the hypersphere. Then, the mean shift motion vector is given by:

$$M(\mathcal{A}) = \frac{\sum_{\mathcal{A}_j \in S_h(\mathcal{A})} \mathcal{A}_j}{n_{\mathcal{A}}} - \mathcal{A}. \tag{15}$$

Once the mean shift filtering procedure is applied, the convergence points are grouped. To cope with outliers in the motion estimation process, one can eliminate spatial regions containing a small number of data. The number of different motion regions and the corresponding affine parameters can then be obtained. Examples of such pre-processing clustering steps are shown in [Figure (2)]. The next step is to decompose the image plane into segments with consistent motion.

3. Reconstruction of Motion Layers

Let us consider a decomposition of the image domain into N motion regions. Such regions are associated with different parametric models $\mathcal{A}_i$ describing the observed apparent motion and can have multiple components. In this section, we formulate the decomposition task within level set representations [16].

3.1. Level Set Formulation

Consider a level set function ϕ for an evolving contour $\partial\mathcal{R}$ with distance transform as embedding function.

$$\phi(x,y) = \begin{cases} D(s,\partial\mathcal{R}) & , \quad s \in \mathcal{R} \\ 0 & , \quad s \in \partial\mathcal{R} \\ -D(s,\partial\mathcal{R}) & , \quad s \in \Omega - \mathcal{R} \end{cases}, \qquad (16)$$

where $\mathcal{R}$ is the region enclosed by the contour and $\Omega - \mathcal{R}$ the background. Using the Dirac $\delta_\alpha(;)$ and the Heaviside distribution $\mathcal{H}_\alpha(\phi)$ with α span as suggested in [23]:

$$\delta_\alpha(\phi) = \begin{cases} 0 & , |\phi| > \alpha \\ \frac{1}{2\alpha}\left(1 + \cos\left(\frac{\pi\phi}{\alpha}\right)\right) & , |\phi| < \alpha \end{cases}, \qquad (17)$$

$$\mathcal{H}_\alpha(\phi) = \begin{cases} 1 & , \phi > \alpha \\ 0 & , \phi < -\alpha \\ \frac{1}{2}\left(1 + \frac{\phi}{\alpha} + \frac{1}{\pi}\sin\left(\frac{\pi\phi}{\alpha}\right)\right) & , |\phi| < \alpha \end{cases}, \qquad (18)$$

a dual image partition can be determined according to $\mathcal{H}_a(\phi)$ and $1 - \mathcal{H}_a(\phi)$. Image partitions with more than two hypotheses can be handled through the use of more level set representations ϕ_i, one per hypothesis i [23]; however, it is not efficient from the computational point of view.

In [21] an efficient tool $-$ driven from the formulation suggested in [23] $-$ was proposed to overcome the above limitations. Without loss of generality and for clarity purposes, let us consider an image partition that requires four distinct classes $(\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \mathcal{R}_4)$. Such partition can be obtained using two level set functions, ϕ_a and ϕ_b, which is equivalent to propagating two contours, as follows:

$$\begin{cases} \mathcal{R}_1 : \phi_a > 0 \;\&\; \phi_b > 0;\ \mathcal{H}_\alpha(\phi_a)\mathcal{H}_\alpha(\phi_b) \\ \mathcal{R}_2 : \phi_a > 0 \;\&\; \phi_b < 0;\ \mathcal{H}_\alpha(\phi_a)(1 - \mathcal{H}_\alpha(\phi_b)) \\ \mathcal{R}_3 : \phi_a < 0 \;\&\; \phi_b > 0;\ (1 - \mathcal{H}_\alpha(\phi_a))\mathcal{H}_\alpha(\phi_b) \\ \mathcal{R}_4 : \phi_a < 0 \;\&\; \phi_b < 0;\ (1 - \mathcal{H}_\alpha(\phi_a))(1 - \mathcal{H}_\alpha(\phi_b)) \end{cases}$$

where $\mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3 \cup \mathcal{R}_4 = \Omega$.

Such formulation can be extended to represent an image partition with N distinct regions $\mathcal{R}_i, i = 1, \cdots, N$, using $\lceil \log_2 N \rceil$ level set representations $\phi_k, k = 1, \cdots, \lceil \log_2 N \rceil$. Each region can be described according to

$$\mathcal{R}_i : \prod_{k=1}^{\lceil \log N \rceil} f(\mathcal{H}_\alpha(\phi_k), i) \qquad (19)$$

where

$$f(\mathcal{H}_\alpha(\phi_k), i) = \begin{cases} \mathcal{H}_\alpha(\phi_k), & \left\lfloor (i-1)/2^{\lceil \log_2 N \rceil - k} \right\rfloor \mod 2 = 0 \\ 1 - \mathcal{H}_\alpha(\phi_k), & \left\lfloor (i-1)/2^{\lceil \log_2 N \rceil - k} \right\rfloor \mod 2 = 1 \end{cases}. \qquad (20)$$

More details on such formulation can be found in [21].

3.2. Smoothness

Natural scenes observed from a camera exhibit smooth boundaries. Such assumption is quite common and can be considered in various forms within the objective function. The most common term, related to the smoothness condition, refers to a minimal curve length/surface area for the propagated contours. Such a constraint can be expressed in the form as:

$$E_{sm}(\phi_1, ..., \phi_{\lceil \log_2 N \rceil}) = \sum_{k=1}^{\lceil \log_2 N \rceil} \iint_\Omega \delta_\alpha(\phi_k)|\nabla \phi_k| \, d\Omega \qquad (21)$$

that aims at minimizing the arc-length of each curve. This term is not equivalent to the one recovered when each hypothesis is represented using a level set function. However, it is a reasonable approximation and guarantees a smooth solution. Calculus of variations can provide a geometric flow to evolve the various level set representations towards the lowest potential of the smoothness term:

$$\frac{d\phi_k}{dt} = -\delta_\alpha(\phi_k)\operatorname{div}\left(\frac{\nabla \phi_k}{|\nabla \phi_k|}\right) \qquad (22)$$

Next, we will introduce image-driven forces that account for smooth visual properties of the scene and explore information in the temporal domain, using the optical flow constraint.

3.3. Visual Grouping

Visual grouping according to color similarity is equivalent to image segmentation. Within our motion reconstruction example we assume that each motion layer has some consistent intensity/visual properties. Then, introducing a module that aims at segmenting according to such properties can improve the spatial segmentation of these layers.

Regional, and global intensity terms are quite popular in image segmentation [24]. The central idea is to recover a metric (region descriptor) that quantifies the fit of a given intensity with the various hypotheses. Parametric approximations (mixture of Gaussian) of the empirical distribution (intensity properties) are not valid for the considered application. The image/motion segments correspond to projections of structures of the real scene with constant depth, which may contain multiple colors and/or texture structures. Therefore, assuming continuous distributions over color space to describe the visual properties of the motion layers is not applicable.

Non-parametric approximation of the empirical distribution is a more realistic assumption for each motion plane. To this end, Parzen windows is a well known technique

$$p_i(I) = \frac{1}{n}\sum_{j=1}^{n} G(I - I_j; \sigma_I), \tag{23}$$

where i refers to the motion layer, n is the number of pixels in the motion layer, I_j is the corresponding intensity, and $G()$ is a one-dimensional zero-mean differentiable Gaussian kernel. These densities can now be used to impose additional constraints on the motion segmentation.

Within a level set formulation, the geodesic active region model [18] is a paradigm for visual grouping. The optimal segmentation corresponds to a maximum posterior grouping probability. Under the assumption that all partitions are equally probable, such partition can be recovered through the lowest potential of:

$$\begin{aligned} &E_{sg}(\phi_1, ..., \phi_{\lceil \log_2 N \rceil}) = \\ &-\sum_{i=1}^{N} \iint_{\Omega} \prod_{k=1}^{\lceil \log_2 N \rceil} f(\mathcal{H}_\alpha(\phi_k), i) \log\left(p_i(I)\right)\ d\Omega, \end{aligned} \tag{24}$$

where $I = I(;\tau)$ and $[p_i,\ i = 1, \cdots, N]$ are the non-parametric approximations of the intensity distribution of the different motion layers. Each integral of the above objective function component measures the quality of fitting between the actual observations and the expected properties of each motion layer. Furthermore, we can recover the lowest potential of objective function with respect to ϕ_k using a gradient descent method:

$$\frac{d\phi_k}{dt} = \sum_{i=1}^{N} g(\mathcal{H}(\phi_k), i) \prod_{j \neq k}^{\lceil \log_2 N \rceil} f(\mathcal{H}(\phi_j), i) \log\left(p_i(I)\right), \tag{25}$$

where

$$\begin{aligned} g(\mathcal{H}_\alpha(\phi_k), i) &= \frac{df(\mathcal{H}_\alpha(\phi_k), i)}{d\phi_k} = \\ &\begin{cases} \delta_\alpha(\phi_k), & \lfloor (i-1)/2^{\lceil \log_2(N) \rceil - k} \rfloor \mod 2 = 0 \\ -\delta_\alpha(\phi_k), & \lfloor (i-1)/2^{\lceil \log_2(N) \rceil - k} \rfloor \mod 2 = 1 \end{cases}. \end{aligned} \tag{26}$$

The interpretation of such a flow is simple. It acts as an adaptive balloon force based on a relative comparison between the probabilities of the two conflicting color distribution hypotheses for the given image color. The non-parametric densities are derived/updated from the latest motion segmentation. The intensity distribution for the image regions that correspond to the different motion layers are updated on-the-fly. The next constraint to be considered is the visual consistency over time.

3.4. Motion Consistency

To this end, one can consider - given the motion models - the optical flow constraint. The objective function for visual consistency for region i with motion parameters $\mathcal{A}_i$ can be formulated as:

$$\begin{aligned} &E(\mathcal{A}_i) = \\ &\iint_{\Omega} \underbrace{\prod_{k=1}^{\lceil \log_2 N \rceil} f(\mathcal{H}_\alpha(\phi_k), i)}_{\mathcal{R}_i\ Hypothesis} \underbrace{\rho(I(;\tau) - I(\mathcal{A}_i; \tau+1))}_{\mathcal{R}_i\ Visual\ Consistency} d\Omega, \end{aligned} \tag{27}$$

where f and ρ have been defined previously, and s is omitted to simplify the notation. Such term is similar to the one considered in [17] for determining the motion of moving objects in sequences with static background.

Given an image partition, the estimation of such motion model can be obtained efficiently using a number of constraints (number of pixels in the $\mathcal{R}_i$ region) that is far superior from the number of unknown variables (number of parameters of the motion model).

Such constraint can be naturally extended to account for all motion layers:

$$\begin{aligned} &E_{mt}(\mathcal{A}_1, ..., \mathcal{A}_N) = \\ &\sum_{i=1}^{N} \iint_{\Omega} \prod_{k=1}^{\lceil \log_2 N \rceil} f(\mathcal{H}_\alpha(\phi_k), i) \rho(I(;\tau) - I(\mathcal{A}_i; \tau+1)) d\Omega, \end{aligned} \tag{28}$$

leading to robust estimation of the motion parameters given the image partition. The calculus of variations and a gradient descent method with respect to $[\phi_k, k = 1, ..., \lceil \log_2 N \rceil]$ and $[\mathcal{A}_i,\ i = 1, ..., N]$ can provide a series of equations to update the partition as well as to recover the optimal motion parameters. The calculus of variations with respect ϕ_k lead to the following flows:

$$\begin{aligned} &\frac{d\phi_k}{dt} = \\ &\sum_{i=1}^{N} g(\mathcal{H}(\phi_k), i) \prod_{j \neq k}^{\lceil \log_2 N \rceil} f(\mathcal{H}_\alpha(\phi_j), i) \rho(I(;\tau) - I(\mathcal{A}_i; \tau+1)), \end{aligned} \tag{29}$$

where g as defined in [Eq. (26)]. The flows that guide the estimation of the motion parameters can be derived in a straightforward manner.

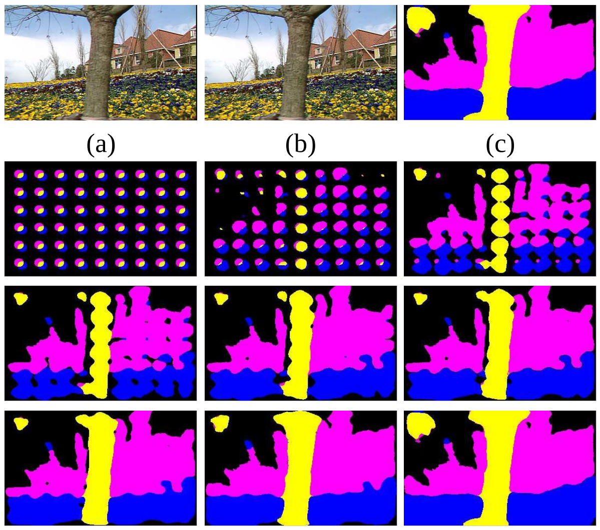


Figure 3: Evolution of Motion Reconstruction for frames $10, 11$ of the Garden Sequence presented in raster-scan format. (a) frame τ, (b) frame $\tau + 1$, (c) motion segmentation.

3.5. Complete Model

Smoothness constraints, visual grouping components and motion/optical flow terms can be integrated to couple motion reconstruction and motion estimation:

$$E(\phi_1, ..., \phi_{\lceil \log_2 N \rceil}, \mathcal{A}_1, ..., \mathcal{A}_N) = E_{sm}(\phi_1, ..., \phi_{\lceil \log_2 N \rceil}) + w_{vc} E_{vc}(\phi_1, ..., \phi_{\lceil \log_2 N \rceil}) + w_{mt} E_{mt}(\mathcal{A}_1, ..., \mathcal{A}_N) \quad (30)$$

where the calculus of variations will provide geometric flows that consist of the different components as earlier presented. The resulting equations aim to evolve an initial partition towards successful separation of the motion layers while estimating the corresponding motion models.

Towards fast implementation of the proposed framework, one can consider the narrow band method [6, 1]. The central idea behind this technique is to evolve the level set representation in the vicinity of the latest position of the contour. Changes on the evolving contour will happen first on the zero-level and then being propagated in the inwards and outwards direction. Some experiment results that do demonstrate the propagation of such implicit representations according the obtained flow are shown in [Figure (3)].

4. Discussion

In this paper we have proposed a variational formulation for the analysis of scenes observed from a moving camera. The objective of such framework was to separate the image plane into various planes with consistent motion parameters. To this end, we have proposed a two-stage approach.

In the first step, the number of motion layers and motion parameters are determined. We adopt a robust technique to estimate the motion parameters using a sliding window technique. These motion parameters are clustered to determine the number of motion layers. Non-parametric clustering $-$ mean shift algorithm $-$ on such high-dimensional space provides an estimate on the number of motion layers and the corresponding motion parameters.

Within the second step, smoothness constraints, motion information and visual grouping modules are integrated to derive the optimal partition (motion layers) and the corresponding parametric motion models. The optical flow constraint in a robust estimation fashion couples the separation of layers and the estimation of their motion parameters. Visual grouping term is considered to improve the performance on uniform data and discontinuities where motion estimation is not accurate. Promising experimental results demonstrate the potentials of the proposed technique [Figure (4,5,6)].

One can claim that the two-stage approach proposed in this paper introduces an important bias in the estimation process. Motion reconstruction will fail to provide optimal results if the number of motion planes is not properly defined. Therefore, a complete framework that aims to determine the number of motion planes on-the-fly is the main future direction of our research.

Future directions of our approach involve integration of information across the temporal domain. The current approach performs the segmentation based on two consequent image frames. Longer-term dynamics can be taken into account for better segmentation when more image frames are considered at the same time. A more appropriate framework can involve the definition of similar concept in 3D [12], where hyper-surfaces are considered to recover and reconstruct the motion layers. Furthermore, proper terms along the region boundaries may better capture the details of each layer.

References

[1] D. Adalsteinsson and J. Sethian. A Fast Level Set Method for Propagating Interfaces. *Journal of Computational Physics*, 118:269–277, 1995.

[2] S. Ayer and H. Sawhney. Layered Representation of Motion Video Using Robust Maximum-Likelihood Estimation of Mixture Models and MDL Encoding. In *IEEE International Conference in Computer Vision*, pages 777–784, Caibridge, USA, 1995.

[3] J. Barron, D. Fleet, and S. Beauchemin. Performance of Optical Flow Techniques. *International Journal of Computer Vision*, 12:43–77, 1994.

[4] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani. Hierarchical Model-Based Motion Estimation. In *European Conference on Computer Vision*, pages 237–252, 1992.

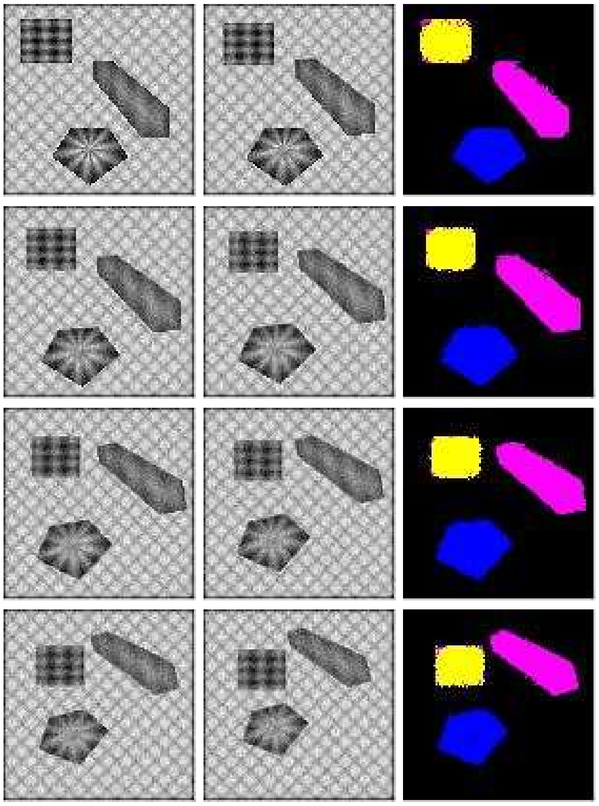


Figure 4: Reconstruction for the synthesized sequence presented in raster-scan format. (a) frame τ, (b) frame $\tau + 1$, (c) motion segmentation.

[5] Y. Cheng. Mean Shift, Mode Seeking, and Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:790–799, 1995.

[6] D. Chopp. Computing Minimal Surfaces via Level Set Curvature Flow. *Journal of Computational Physics*, 106:77–91, 1993.

[7] D. Comaniciu and P. Meer. Mean Shift: A Robust Approach Toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:603–619, 2002.

[8] Trevor Darrell and Alex Pentland. Cooperative Robust Estimation Using Layers of Support. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):474–487, 1995.

[9] B. Horn and B. Schunck. Determinating Optical Flow. *Artificial Intelligence*, 17:185–203, 1981.

[10] P. Huber. *Robust Statistics*. John Wiley & Sons, 1981.

[11] M. Isard and A. Blake. Contour Tracking by Stochastic Propagation of Conditional Density. In *European Conference on Computer Vision*, volume I, pages 343–356, 1996.

[12] A. Mansouri and J. Konrad. Multiple motion segmentation with level sets. *IEEE Transactions on Image Processing*, 24:201–220, 2003.

[13] P. Meer, D. Mintz, D.Y. Kim, and A. Rosenfeld. Robust Regression Methods for Computer Vision: A Review. *International Journal of Computer Vision*, 6:59–70, 1991.

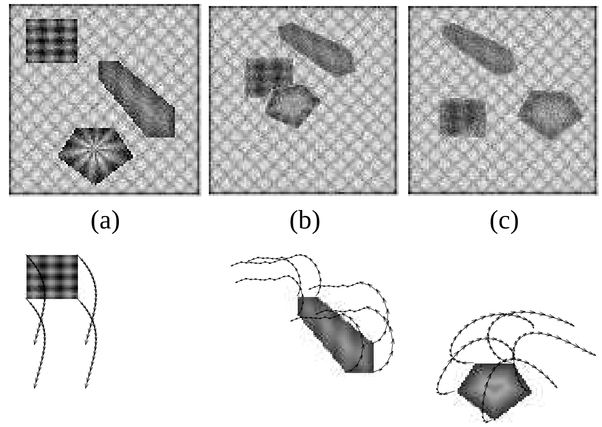


Figure 5: Recovered motion for synthesized sequence. (a)-(c) Starting, middle, and ending frames in the sequence , (d)-(f) the resulting motion for different objects during the whole sequence

[14] J-M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6:348–365, 1995.

[15] S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer. 2003.

[16] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed : Algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[17] N. Paragios and R. Deriche. Geodesic Active Regions for Motion Estimation and Tracking. In *IEEE International Conference in Computer Vision*, pages 688–674, 1999.

[18] N. Paragios and R. Deriche. Geodesic Active Regions: A New Framework to Deal with Frame Partition Problems in Computer Vision. *Journal of Visual Communication and Image Representation*, 13:249–268, 2002.

[19] C. Stauffer and W. Grimson. Adaptive Background Mixture Models for Real-time Tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages II:246–252, Colorado, USA, 1999.

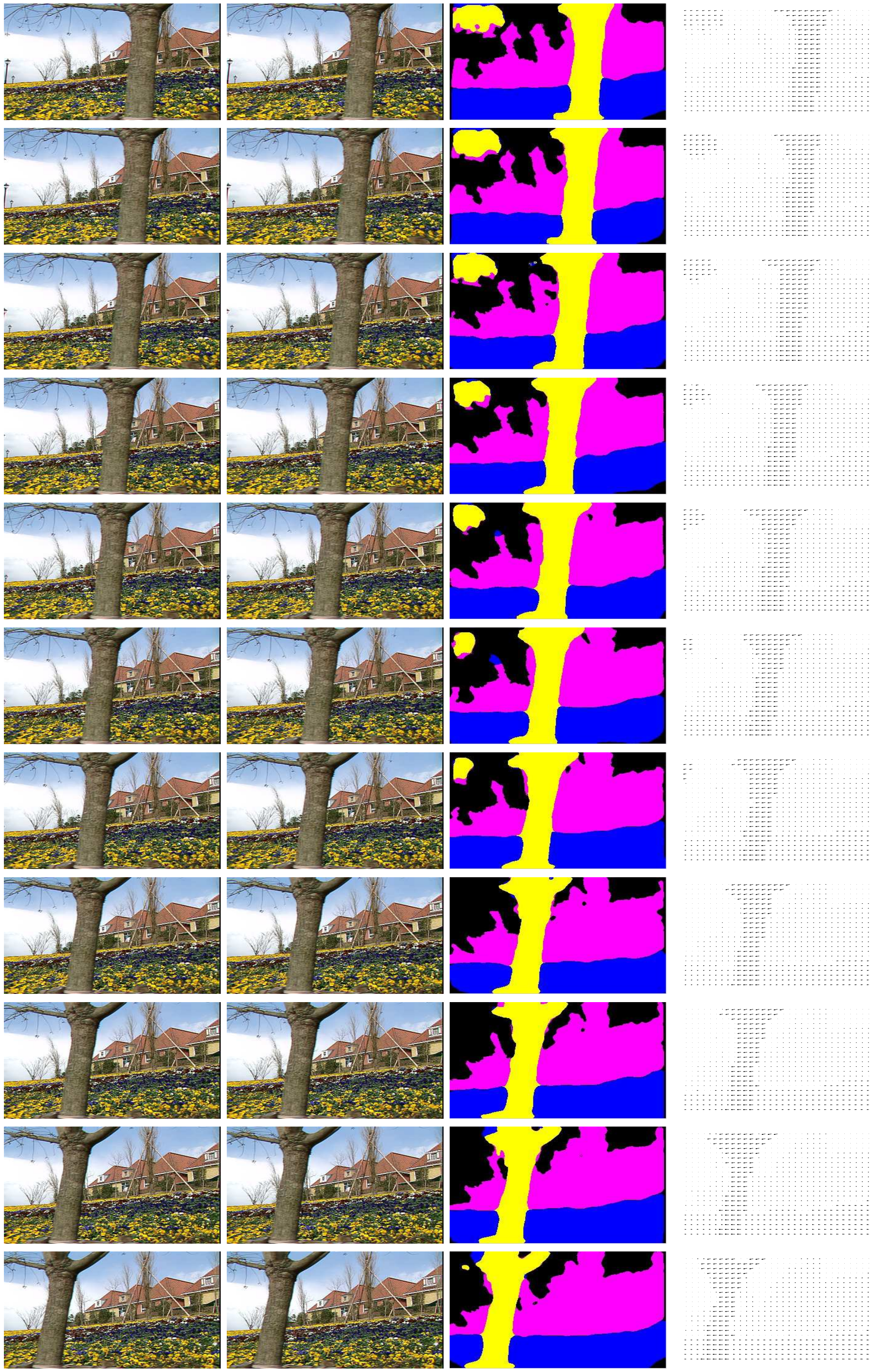
[20] A. Tikhonov. *Ill-Posed Problems in Natural Sciences*. Coronet, 1992.

[21] L. Vese and T. Chan. A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model. *International Journal of Computer Vision*, 50:271–293, 2002.

[22] J. Wang and E. Adelson. Representing Moving Images with Layers. *IEEE Transactions on Image Processing*, 3(5):625–638, 1994.

[23] H-K. Zhao, T. Chan, B. Merriman, and S. Osher. A variational Level Set Approach to Multiphase Motion. *Journal of Computational Physics*, 127:179–195, 1996.

[24] S. Zhu and A. Yuille. Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:884–900, 1996.



(a) (b) (c) (d)

Figure 6: Reconstruction for the garden Sequence presented in raster-scan format (Part 1). (a) frame τ, (b) frame $\tau + 1$, (c) motion segmentation, (d) Optical flow estimation for the Garden sequence in raster-scan format.