Spatio-temporal motion segmentation via level set partial differential equations

Abdol-Reza Mansouri, Amar Mitiche, and Rosario El-Feghali
INRS Telecommunications, Place Bonaventure, 900 de la Gauchetiere Ouest, C.P. 644, Montreal, Quebec H5A 1C6, Canada

Abstract

Motion-based segmentation of image sequences is an important problem of image analysis, with numerous applications to image coding and image manipulation. We present a novel algorithm for segmenting image sequences into distinct motion trajectories. This algorithm is expressed as a system of level set partial differential equations which is solved using level set discretization schemes. We provide experimental results on real image sequences with natural and synthetic image motion.

1 Introduction

Segmentation of an image sequence into moving regions is one of the most important and difficult problems in video processing (object-based video database querying, object-based frame conversion), video compression (object-based coding, e.g., MPEG4) and computer vision (scene analysis and interpretation). In this paper, we propose a novel approach to motion segmentation by formulating it as the computation of the volume spanned by moving regions in 3-dimensional spatio-temporal space. This is in sharp contrast to standard approaches which treat motion segmentation as a sequence of purely spatial segmentations. Spatio-temporal motion segmentation has direct applications to volume coding, allowing the full use of spatial and temporal correlations in image sequences, leading to potentially higher compression rates.

We formulate the problem of motion segmentation in spatio-temporal space as a Bayesian estimation problem, and we show that this in turn leads to the minimization of a particular energy functional of the segmentation. The Euler-Lagrange descent equation which minimizes this functional is expressed as a level set partial differential equation, and is then generalized to the multiple motion case, yielding a coupled system of level set partial differential equations. We illustrate our proposed algorithm on a real image sequences with both natural and synthetic motion. The algorithm presented here generalizes our previous work [1] on motion-based image segmentation to motion-based *image sequence* segmentation via level set partial differential equations.

2 Spatio-temporal Motion Segmentation

2.1 Basic Models

Let I be an image sequence function defined over a domain $\boldsymbol{D} = \Omega \times [0,T]$, where Ω is an open subset of $\mathbb{R}^2$. I is thus a map $I : \boldsymbol{D} \to \mathbb{R}, \quad (\boldsymbol{x},t) \mapsto I(\boldsymbol{x},t)$. For simplicity of presentation, we shall first consider the case of a moving object against a still or moving background. We thus assume that there exist subsets $\Omega_1, \Omega_2 \subset \Omega$ and continuous functions $f_1, f_2 : \Omega \times [0,T] \to \Omega$ such that $\Omega_1 \cap \Omega_2 = \emptyset$ and $\Omega_1 \cup \Omega_2 = \Omega$, and $\forall t \in [0,T]$:

$$\begin{cases} I(f_1(\boldsymbol{x},t),t) = I(\boldsymbol{x},0) + \sigma B_{\boldsymbol{x}}(t), \forall \boldsymbol{x} \in \Omega_1, \\ I(f_2(\boldsymbol{x},t),t) = I(\boldsymbol{x},0) + \sigma B_{\boldsymbol{x}}(t), \forall \boldsymbol{x} \in \Omega_2, \end{cases}$$

where $B_{\boldsymbol{x}}$ is standard Brownian motion [2] for all $\boldsymbol{x} \in \Omega$ and where $B_{\boldsymbol{x}}$ and $B_{\boldsymbol{y}}$ are independent whenever $\boldsymbol{x} \neq \boldsymbol{y}$. We define $\boldsymbol{D}_1, \boldsymbol{D}_2 \subset \boldsymbol{D}$ by:

$$\boldsymbol{D}_1 = \bigcup_{t \in [0,T]} f_1(\Omega_1, t), \quad \boldsymbol{D}_2 = \bigcup_{t \in [0,T]} f_2(\Omega_2, t),$$

and we assume $\boldsymbol{D}_1, \boldsymbol{D}_2$ form a partition of $\boldsymbol{D}$. $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are the volumes comprising the motion trajectories of motion regions Ω_1 and Ω_2, respectively, or, equivalently, the volumes spanned by these motion regions throughout the sequence. The problem of spatio-temporal motion segmentation can then be defined as that of estimating $\boldsymbol{D}_1, \boldsymbol{D}_2 = \boldsymbol{D}_1^c, f_1, f_2$ given the image sequence I. This is in sharp contrast to motion-based *image* segmentation algorithms, which would estimate Ω_1 and Ω_2, as opposed to $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$.

2.2 MAP Estimation

The maximum a posteriori estimate of $\boldsymbol{D}_1, f_1, f_2$ given I is given by:

$$\begin{aligned}(\boldsymbol{D}_1^\star, f_1^\star, f_2^\star) &= \arg\max_{(\boldsymbol{D}_1, f_1, f_2)} P(\boldsymbol{D}_1, f_1, f_2 | I) \\ &= \arg\max_{(\boldsymbol{D}_1, f_1, f_2)} P(I|\boldsymbol{D}_1, f_1, f_2) P(\boldsymbol{D}_1, f_1, f_2) \qquad (1)\end{aligned}$$

The first term on the product in the right-hand side of equation (1) is the likelihood of observing the image sequence I given the segmentation $\{\boldsymbol{D}_1, \boldsymbol{D}_1^c\}$ of the spatio-temporal image sequence domain $\boldsymbol{D}$ and the motion trajectory functions f_1 and f_2, while the second term is the prior on $\boldsymbol{D}_1$ and on the motion trajectory functions. In all that follows, we shall simplify the problem by estimating f_1 and f_2 prior to the segmentation; we shall then estimate $\boldsymbol{D}_1$ using the trajectory function estimates $f_1^\star$ and $f_2^\star$ through the maximum a posteriori estimate:

$$\boldsymbol{D}_1^\star = \arg\max_{\boldsymbol{D}_1} P(I|\boldsymbol{D}_1, f_1^\star, f_2^\star) P(\boldsymbol{D}_1) \qquad (2)$$

where we have assumed the statistical independence of the segmentation and the trajectory functions. We choose a prior on $\boldsymbol{D}_1$ that favors minimum area of the boundary $\partial\boldsymbol{D}_1$; this is by no means the only choice of prior that can be made, and in case a priori information is known about the geometry of the segmentation $\{\boldsymbol{D}_1, \boldsymbol{D}_1^c\}$, it can be incorporated as a prior in the term $P(\boldsymbol{D}_1)$ of equation (2). Up to an additive constant, which can be neglected without loss of generality, we then have:

$$-\log P(\boldsymbol{D}_1) = \int_{\partial\boldsymbol{D}_1} d\rho,$$

where $d\rho$ is the surface measure on $\partial\boldsymbol{D}_1$. Consider now an image sequence $\hat{I}$ which is defined over the same spatio-temporal domain as the image sequence I. If $\hat{I}$ and I differ by an additive constant, i.e. $\hat{I} = I + c$, for some $c \in \mathbb{R}$, then the motion-based segmentation of $\hat{I}$ should be identical to that of I. If now $\hat{I}$ and I differ by an additive function which is constant along the motion trajectories of I, then the motion trajectories of $\hat{I}$ and I are identical, and hence, $\hat{I}$ and I yield identical motion-based segmentations. Combining these two observations, we have:

$$\boldsymbol{D}_1^\star = \arg\max_{\boldsymbol{D}_1} P(d(I \circ \gamma)_{\gamma\in\Lambda} | \boldsymbol{D}_1, f_1^\star, f_2^\star) P(\boldsymbol{D}_1),$$

where Λ is the set of all motion trajectories of I, and d is the differential operator. Since for $\gamma \in \Lambda$ $I \circ \gamma$ is a Brownian process, and since for two distinct trajectories in Λ, these processes are statistically independent, the MAP estimation problem in (2) is equivalent to the following energy minimization problem:

$$\vec{\Gamma}^\star = \arg\min_{\vec{\Gamma}} E(\vec{\Gamma}),$$

where the energy function E is given by:

$$\begin{aligned}E(\vec{\Gamma}) &= \frac{1}{2\sigma^2}\int_{\boldsymbol{z}\in\boldsymbol{D}_{\vec{\Gamma}}} \Psi_1(\boldsymbol{z})d\boldsymbol{z} \\ &+ \frac{1}{2\sigma^2}\int_{\boldsymbol{z}\in\boldsymbol{D}_{\vec{\Gamma}}^c} \Psi_2(\boldsymbol{z})d\boldsymbol{z} \\ &+ \lambda\int_{\vec{\Gamma}} d\rho,\end{aligned}$$

and where $\Psi_i : \boldsymbol{D} \to \mathbb{R}$ $(i = 1, 2)$ denotes the function

$$\boldsymbol{z} \mapsto \Psi_i(\boldsymbol{z}) = (< \vec{\nabla} I, \frac{df_i^\star}{dt} > (\boldsymbol{z}) + \frac{\partial I}{\partial t}(\boldsymbol{z}))^2,$$

$\vec{\Gamma}$ is a closed surface that is the estimator of the boundary $\partial\boldsymbol{D}_1$ of $\boldsymbol{D}_1$, $\boldsymbol{D}_{\vec{\Gamma}}$ is the spatio-temporal domain enclosed by $\vec{\Gamma}$, $\boldsymbol{D}_{\vec{\Gamma}}^c$ is its complement, $d\rho$ is the surface measure on $\vec{\Gamma}$, and λ a penalty coefficient on the area of the surface $\vec{\Gamma}$. It is important to note that the functions Ψ_i are nothing other than motion compensated residuals, for the particular motion trajectory functions $f_i^\star$, $i = 1, 2$.

2.3 Surface Evolution Equations

The minimization of the energy functional E with respect to the surface $\vec{\Gamma}$ can be achieved by considering $\vec{\Gamma}$ as being also a function of (algorithmic) time τ. In order for $\lim_{\tau\to\infty}\vec{\Gamma}(\tau)$ to be a (local) minimum of E, $\tau \mapsto \vec{\Gamma}(\tau)$ has to satisfy the following differential equation:

$$\frac{d\vec{\Gamma}}{d\tau} = -\frac{\delta E}{\delta\vec{\Gamma}}, \qquad (3)$$

where $\frac{\delta E}{\delta\vec{\Gamma}}$ is the functional derivative of E with respect to $\vec{\Gamma}$. Indeed, we then obtain:

$$\begin{aligned}\frac{d(E\circ\vec{\Gamma})}{d\tau} &= < \frac{d\vec{\Gamma}}{d\tau}, \frac{\delta E}{\delta\vec{\Gamma}} > \\ &= - < \frac{\delta E}{\delta\vec{\Gamma}}, \frac{\delta E}{\delta\vec{\Gamma}} > \\ &\leq 0,\end{aligned}$$

and $\tau \mapsto E \circ \vec{\Gamma}(\tau)$ is thus a decreasing function of τ. Writing in more detail the Euler-Lagrange descent equation (3), we obtain that $\tau \mapsto \vec{\Gamma}(\tau)$ should satisfy the following differential equation:

$$\frac{d\vec{\Gamma}}{d\tau}(\boldsymbol{p},\tau) = -[\frac{1}{2\sigma^2}(\Psi_1(\boldsymbol{p}) - \Psi_2(\boldsymbol{p})) + \lambda H(\boldsymbol{p},\tau)]\vec{N}(\boldsymbol{p},\tau)$$

where $\boldsymbol{p}$ is a point on surface $\vec{\Gamma}$, H is the mean curvature of $\vec{\Gamma}$ and $\vec{N}$ is the outward unit normal to $\vec{\Gamma}$.

2.4 Level Set Evolution Equations

We can choose to numerically solve the Euler-Lagrange descent equation (3) *explicitely*, by discretizing the surface $\vec{\Gamma}$ using a fixed number of nodes. Such an approach can lead to large approximation errors, especially in the computation of derivatives of higher order (as with curvature), depending on the evolution of the surface, since discretization errors vary according to the evolution of the surface. A better numerical approach is to represent the surface $\vec{\Gamma}$ *implicitly* by the zero level set of a function $u : \boldsymbol{D} \rightarrow \mathbb{R}$. We adopt the convention that the "inside" $\boldsymbol{D}_{\vec{\Gamma}}$ of the surface $\vec{\Gamma}$ will correspond to the set $\{u > 0\}$ and the "outside" $\boldsymbol{D}^c_{\vec{\Gamma}}$ to the set $\{u < 0\}$. The surface itself will thus be the set of all points in $\boldsymbol{D}$ which are in the zero level set of u, that is $\vec{\Gamma} = \{\boldsymbol{z} \in \boldsymbol{D} | u(\boldsymbol{z}) = 0\}$. There are two major benefits to this implicit representation of $\vec{\Gamma}$: numerical stability and topology independence. Indeed, discretization problems that occur with an explicit representation of $\vec{\Gamma}$ do not occur with level sets since each level set function is defined over a discrete grid with uniform spacing, and all derivatives are computed using finite difference approximations. Also, a surface evolution equation where the surface is explicitely represented does not allow for changes in surface topology, and a simply connected surface will remain simply connected throughout its evolution. On the other hand, surfaces defined implicitly using level set functions can freely change topology during their evolution; for example, a sphere can evolve into a torus, and vice-versa. Viewing u also as a function of τ, it is a simple matter to show that for the zero level set of u to follow the evolution of $\vec{\Gamma}$, u itself should evolve according to the following partial differential equation [3]:

$$\frac{\partial u}{\partial \tau}(\boldsymbol{z}, \tau) = -[\frac{1}{2\sigma^2}(\Psi_1(\boldsymbol{z}) - \Psi_2(\boldsymbol{z})) + \lambda H_u(\boldsymbol{z}, \tau)]\|\vec{\nabla} u\|$$

where the mean curvature H_u is given by $H_u = \vec{\nabla} \cdot \frac{\vec{\nabla} u}{\|\vec{\nabla} u\|}$. The behaviour of the level set function u can be again read directly from this evolution equation. Without the curvature term, if a point $\boldsymbol{z}$ is in $\boldsymbol{D}_1$, then the difference of the motion residuals $\Psi_1(\boldsymbol{z}) - \Psi_2(\boldsymbol{z})$ will be negative, and the partial derivative $\frac{\partial u}{\partial \tau}(\boldsymbol{z})$ will be positive, forcing $u(\boldsymbol{z})$ to increase, and eventually become positive at $\boldsymbol{z}$. If on the other hand $\boldsymbol{z} \in \boldsymbol{D}_2$, then the difference of the motion residuals $\Psi_1(\boldsymbol{z}) - \Psi_2(\boldsymbol{z})$ will be positive, and the partial derivative $\frac{\partial u}{\partial \tau}(\boldsymbol{z})$ will be negative, forcing $u(\boldsymbol{z})$ to decrease, and eventually become negative at $\boldsymbol{z}$. Inclusion of the mean curvature term on the right hand side of the equation above yields diffusion along the level sets of u that locally straightens out the level sets, and globally reduces their area.

2.5 Extension to Multiple Motion Regions

The level set evolution equation obtained in Section (2.4) can be easily generalized to the case of multiple motion regions. We thus assume there exists a partition of the image domain Ω into a family $\{\Omega_i\}_{i=1}^N$ of subsets, and that there exist N distinct trajectory functions $f_i : \Omega \times [0, T] \rightarrow \Omega$ such that:

$$I(f_i(\boldsymbol{x}, t), t) = I(\boldsymbol{x}, 0) + \sigma B_{\boldsymbol{x}}(t),$$

$\forall (\boldsymbol{x}, t) \in \Omega_i \times [0, T]$. We define $\boldsymbol{D}_i \in \boldsymbol{D}$ by

$$\boldsymbol{D}_i = \bigcup_{t \in [0,T]} f_i(\Omega_i, t),$$

and we assume $\{\boldsymbol{D}_i\}_{i=1}^N$ forms a partition of $\boldsymbol{D}$ as well. Defining the family $\{u_i : \boldsymbol{D} \times \mathbb{R}^+ \rightarrow \mathbb{R}\}_{i=1}^N$ of level set functions, with $\boldsymbol{D}_i$ being estimated by the subset $\{u_i > 0\}$, we obtain the following *system* of coupled partial differential equations:

$$\begin{aligned} \frac{\partial u_i(\boldsymbol{z}, \tau)}{\partial \tau} &= -\Big([\Psi_i(\boldsymbol{z}) - \min_{k \neq i, u_k(\boldsymbol{z}, \tau) > 0} \Psi_k(\boldsymbol{z})] \\ &+ \lambda H_{u_i}(\boldsymbol{z}, \tau)\Big)\|\vec{\nabla} u_i(\boldsymbol{z}, \tau)\|, i = 1, \ldots, N. \end{aligned}$$

2.6 Trajectory Models

It is important to note that the trajectory functions f_i appear in the level set evolution equations only through their time derivative $\frac{df_i}{dt}$. To model the various motion trajectories therefore, it is sufficient to describe the differential equation that they have to satisfy. The general form of such a differential equation is given by:

$$\frac{df_i}{dt} = F_i(f_i, t), \quad i = 1, \ldots, N,$$

where $F_i : \boldsymbol{D} \rightarrow \mathbb{R}^2$. In all the experiments described in this paper, we will consider the simplest case of this differential equation, namely

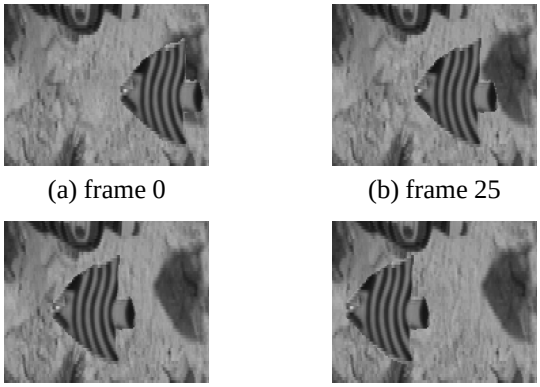
$$\frac{df_i}{dt} = \vec{v}_i, \quad i = 1, \ldots, N,$$

where $\vec{v}_i$ is a constant (velocity) vector. Such a differential equation models trajectories associated with constant velocity translations. Further generalization is afforded by considering constant acceleration trajectories, which satisfy the following differential equation:

$$\frac{df_i}{dt} = \vec{a}_i t + \vec{v}_i, \quad i = 1, \ldots, N.$$

3 Experimental Results

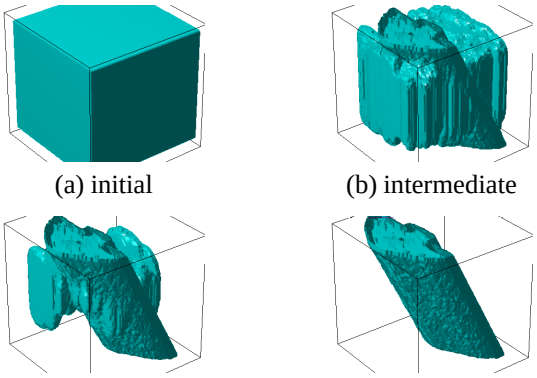
Fig. 1 shows frames 0, 25, 50, and 75 of a sequence with natural texture but synthetic motion constructed from the *aqua* sequence. The striped fish is moving leftwards on a fixed background, with a translational displacement of 1 pixel per frame. Fig. 2 shows the spatio-temporal surface which englobes the various motion trajectories at four instants during its evolution. As can be seen in Fig. 2(d), a cylinder-like region corresponding to the fish has been isolated. Time slices of this spatio-temporal surface corresponding to frames 0, 25, 50, and 75 of the sequence are shown in Fig. 3; Note how precisely the fish has been segmented from the background. Fig. 4 shows frames 0, 5, 10, and 15 of a real outdoors sequence where the two people shown are moving downwards and rightwards, with approximately constant velocity. Fig. 5 shows four instants during the evolution of the spatio-temporal surface englobing the various motion trajectories. Note how the topology of the evolving surface changes, from connected to non-connected. Note also that whereas the initial surface is simply-connected with no boundary, none of the connected components of the final surface is simply connected, and one of them has a boundary. Such topological changes could not have occurred had the evolving surface been represented explicitly. Time slices of this spatio-temporal surface corresponding to frames 0, 5, 10, and 15 of the sequence are shown in Fig. 6. Fig. 7 shows frames 0, 7, 14, and 21 of a real outdoors sequence with two cars moving leftwards with identical almost constant velocity. Fig. 8 shows four instants during the evolution of the spatio-temporal surface englobing the various motion trajectories. Here again, the evolving surface undergoes changes in topology. Time slices of this spatio-temporal surface corresponding to frames 0, 7, 14, and 21 of the sequence are shown in Fig. 9. Note that the cars have been accurately segmented from the background; note also that a portion of the ground comprising the car's shadow has been included in the segmentation, in conformity with the observation model adopted in this algorithm.



(a) frame 0 (b) frame 25

(c) frame 50 (d) frame 75

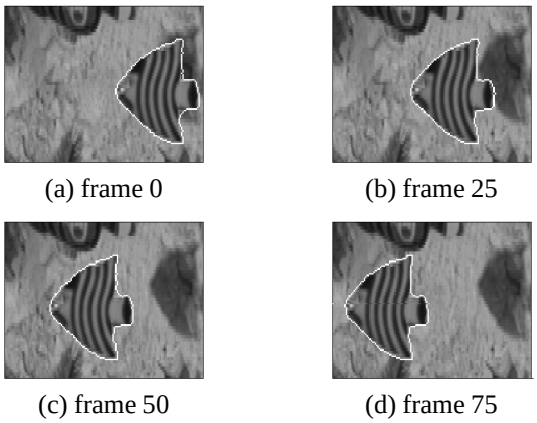
Figure 1. Original sequence.



(a) initial (b) intermediate

(c) intermediate (d) final

Figure 2. Spatio-temporal surface evolution (positive time axis is upwards).



(a) frame 0 (b) frame 25

(c) frame 50 (d) frame 75

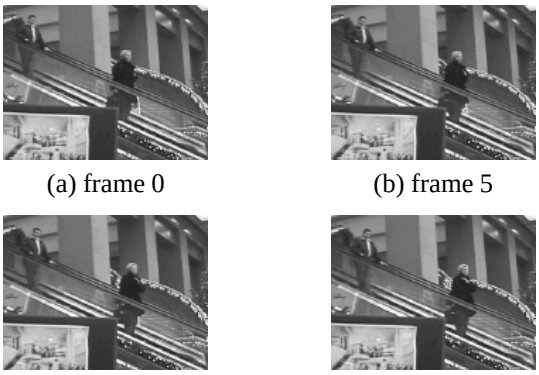
Figure 3. Time slices of spatio-temporal surface.

References

[1] A.-R. Mansouri, J. Konrad, "Multiple motion segmentation with level sets," submitted to *IEEE Transactions on Image Processing*, April 2000 (in review).

[2] B. Oksendal, *Stochastic Differential Equations: An Introduction with Applications*, Springer-Verlag, New-York 1996.

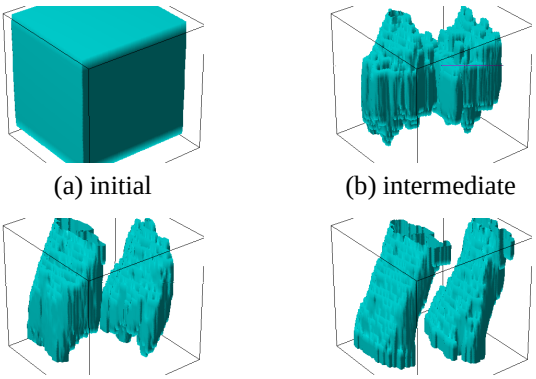
[3] J.A. Sethian, *Level Set Methods*, Cambridge University Press, 1996.



(a) frame 0 (b) frame 5

(c) frame 10 (d) frame 15

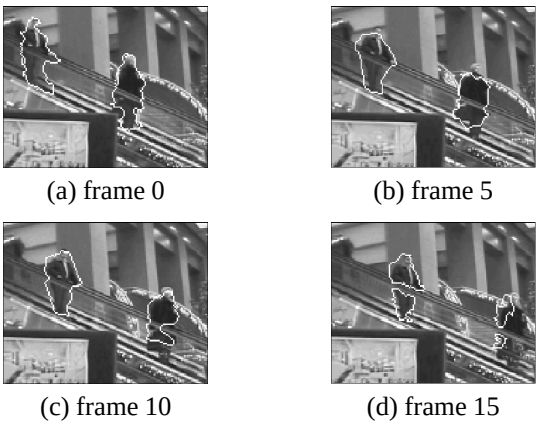
Figure 4. Original sequence.



(a) initial (b) intermediate

(c) intermediate (d) final

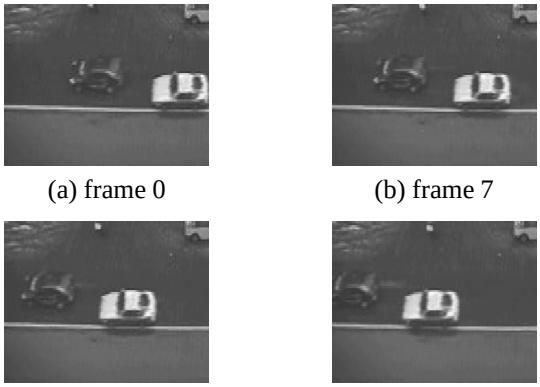
Figure 5. Spatio-temporal surface evolution (positive time axis is upwards).



(a) frame 0 (b) frame 5

(c) frame 10 (d) frame 15

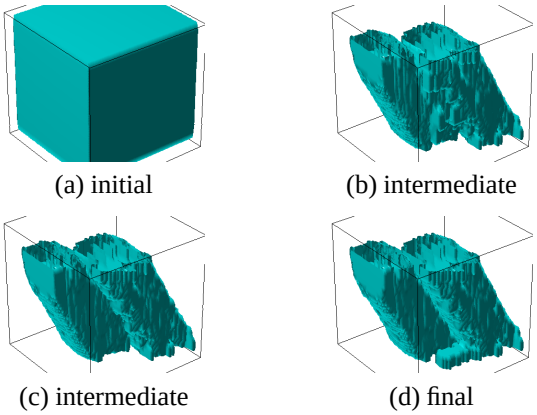
Figure 6. Time slices of spatio-temporal surface.



(a) frame 0 (b) frame 7

(c) frame 14 (d) frame 21

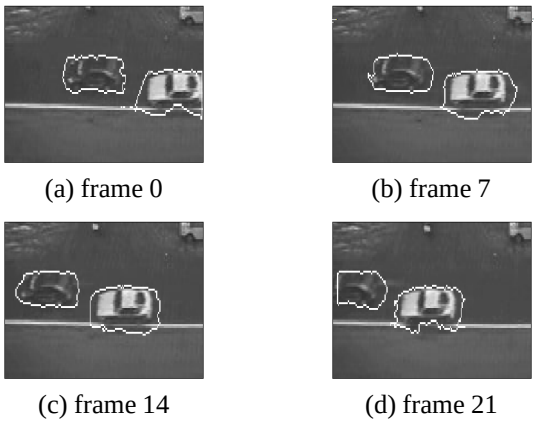
Figure 7. Original sequence.



(a) initial (b) intermediate

(c) intermediate (d) final

Figure 8. Spatio-temporal surface evolution (positive time axis is upwards).



(a) frame 0 (b) frame 7

(c) frame 14 (d) frame 21

Figure 9. Time slices of spatio-temporal surface.