International Journal of Computer Applications (0975 - 8887)
Volume 55 - No. 03, October 2012

Video Segmentation using 2D+time Mumford-Shah Functional

Mohamed El Aallaoui
Laboratory of Mathematical Engineering (LINMA)
Faculty of Sciences -Eljadida- Morocco

Abdelwahad Gourch
Faculté des Sciences Juridiques Économiques et Sociales
Ain Sebaâ -Casablanca- Morocco

ABSTRACT

this paper describes a new video segmentation method obtained by minimizing an extension of Mumford-Shah functional used for 2D+time partitions. This extension permits to write the Mumford-Shah functional as an amultiscale energy, which is minimized on a 2D+time persistent hierarchy. The building of this hierarchy based on connected components of spatio-temporal regions.

Keywords:

Video segmentation, 2D+time Mumford-Shah functional, amultiscale energy, hierarchy, 2D-shapes.

1. INTRODUCTION

Image segmentation is intended to group perceptually similar pixels into 2D regions, and the corresponding border is gained at the same time. Video segmentation generalizes this concept to the grouping of pixels into spatio-temporal regions that exhibit coherence in both appearance and motion, but this generalization pose the complexity of spatio-temporal grouping, and in order to overcome this complexity, the existing video segmentation methods use two Techniques; frame-by-frame and volumetric clustering (3D). Frame-by-frame Techniques filter each frame as an isolated image [1, 2]. Intuitively, these techniques under-exploit the available information and then create associations between regions over time to identify sporadic regions [3, 4, 5]. Although this filtering improves stability, temporal coherence is not ensured because the region map for each frame is formed independently without knowledge of the adjacent frames. Volumetric approaches cluster pixels in 3D space (x, y, t), using unsupervised clustering techniques to group space-time pixels [6, 7, 8], such as mean-shift [9, 10], multi-label propagation [11], or gaussian mixture models [12]. Consequently, these approches treat the temporal coherence as spatial coherence which can not always enforce the consistency of region boundaries over time, and forms disconnected space-time volumes in small or fast moving objects.

This paper presents an efficient and scalable method for spatio-temporal segmentation obtained by minimizing a 2D+time extention of the simplified Mumford-Shah functional. This extension permits to write the Mumford-Shah functional as an amultiscale energy, and using the theory of optimization of amultiscale energy on a hierarchy, we compute a video segmentation by selecting a partition of video domain, which minimizes the amultiscale energy of 2D+time Mumford-Shah Functional on a 2D+time hierarchy.

The outline of the paper is as follows. In section 2, we extend the simplified Mumford-Shah functional for 2D+time video segmentation, and we transform the 2D+time Mumford-Shah segmentation problem to an optimization problem of 2D+time affine energy. The theory of optimization of affine energy on a hierarchy is described in section 3. In section 4 we give a sufficient condition to guarantee that the 2D+time affine energy is amultiscale energy and we present our video segmentation algorithm on 2D+time hierarchy. The building of 2D+time hierarchy is discussed in section 5, and we show experimental results in section 6. Finally, the paper closes with some conclusions in section 7.

2. EXTENSION OF MUMFORD-SHAH FUNCTIONAL FOR 2D+TIME SEGMENTATION

The Mumford Shah functional was introducted by Mumford and Shah in 1989 [13]. It follows:

$$MS_\lambda(\tilde{f}, \mathcal{C}) = \lambda \mathcal{H}(\mathcal{C}) + \int_{\Omega - \mathcal{C}} (\tilde{f}(x) - f(x))^2 dx. \quad (1)$$

Like before, f is our image function. We have $\Omega = \Omega_1 \cup \Omega_2 \cup ... \cup \Omega_n \cup \mathcal{C}$ in which Ω is the domain of our image, Ω_i is the region in our image that represents a section $\mathcal{O}_i$ which does not including the boundaries, and $\mathcal{C}$ is the set of smooth arcs that make up boundaries for the Ω_i. $\mathcal{H}(\mathcal{C})$ denotes the length of the system of curves $\mathcal{C}$. The function $\tilde{f}$ is a piecewise constant image, i.e., constant on each region of $\Omega - \mathcal{C}$, and $\lambda > 0$ is a parameter. Mumford-Shah segmentation of image f is defined by a pair $(\mathcal{C}, \tilde{f})$, witch minimize the Mumford Shah functional $MS_\lambda(\tilde{f}, \mathcal{C})$ [14].

Let $f : \Omega \times [T_i, T_f] \to \mathbb{R}$ be a video sequence with spatial domain Ω and temporal interval $[T_i, T_f]$. We shall assume that the time is discrete $\{t_n\}_{n\in[1,N]}$. Our goal is to compute a segmentation of video sequence $f(\vec{x}, t_n)$ defined by a pair $(\mathcal{C}, \tilde{f})$, such that:

- $\tilde{f}$ is piecewise regular function in $(\Omega \times [T_i, T_f]) - \mathcal{C}$:

$$\begin{aligned} \tilde{f} : \ (\Omega \times [T_i, T_f]) - \mathcal{C} &\longrightarrow \mathbb{R} \\ (\vec{x}, t_n) &\longmapsto \sum_{v_m \in V} \tilde{f}_m \chi_{v_m}(\vec{x}, t_n), \end{aligned}$$

  where V is a 2D+time partition of $\Omega \times [T_i, T_f]$.

- $\mathcal{C}$ is the set of boundaries where $\tilde{f}$ is discontinuous.

Notice that, the set of boundaries $\mathcal{C}$ represents the 2D+time partition $V = \{v_m\}_{m\in[1,M]}$, where v_m is a 2D+time section, which does not including the boundaries, and defined by the 2D section $\mathcal{O}_{n,m}$ at each time t_n,

$$v_m = \bigcup_{n=1}^{N} \mathcal{O}_{n,m}. \quad (2)$$

REMARK 2.1. *At each time t_n, the set $\{\mathcal{O}_{n,m}\}_{m\in[1,M]}$ is a partition of Ω, that generates a set of boundaries. We note $\mathcal{C}_n$ this set of boundaries.*

Hence, the $\mathcal{C}$ is defined by the sets C_n,

$$\mathcal{C} = \bigcup_{n=1}^{N} \mathcal{C}_n. \tag{3}$$

We extend 2D simplified Mumford-Shah functional (1) for computing a segmentation of video sequence defined by a pair $(\mathcal{C}, \tilde{f})$, and we propose the following model:

$$\begin{aligned} E_\lambda(\tilde{f},\mathcal{C}) \;=\; & \lambda\sum_{n=1}^{N}\mathcal{H}(\mathcal{C}_n) + \sum_{n=1}^{N}\int_{\Omega-\mathcal{C}_n}\Big(\tilde{f}(\vec{x},t_n) - f(\vec{x},t_n)\Big)^2 d\vec{x} \\ & + \sum_{n=1}^{N}\sum_{m=1}^{M}\int_{\Omega}\Big(f(\vec{x},t_n)\chi_{\mathcal{O}_{n,m}}(\vec{x}) \\ & - f(\phi_n(\vec{x}),t_{n+1})\chi_{\mathcal{O}_{n+1,m}}(\phi_n(\vec{x}))\Big)^2 d\vec{x}, \end{aligned}$$

where, $\lambda \in \mathbb{R}^+$, $\mathcal{H}(\mathcal{C}_n)$ is measure of $\mathcal{C}_n$ and $\phi_n(\vec{x})$ represents the trajectory of the particle which was in the position $\vec{x}$ at time t_n, and corresponds t_{n+1}. We modeliez the trajectory ϕ_n with an affine model defined by:

$$\phi_n(x_1,x_2) = (u_{n+1}\delta_n t + x_1, v_{n+1}\delta_n t + x_2), \tag{4}$$

Where $\delta_n t = t_{n+1} - t_n$, and (u_{n+1}, v_{n+1}) is the components of optical flow in the horizontal and vertical direction respectively at time t_{n+1} and location (x_1, x_2).

REMARK 2.2. *We observe that, given $\mathcal{C}$, the minimum of $E_\lambda(\tilde{f},\mathcal{C})$ with respect to the variable $\tilde{f}_m$ is explicitly given by*

$$\tilde{f}_m = \frac{1}{\sum_{n=1}^{N}|\mathcal{O}_{n,m}|}\sum_{n=1}^{N}\int_{\mathcal{O}_{n,m}} f(\vec{x},t_n)d\vec{x}. \tag{5}$$

This observation permits to write our model of 2D+time Mumford-Shah functional $E_\lambda(\tilde{f},\mathcal{C})$, as a 2D+time affine energy of $\mathcal{C}$ and denote it by $E_\lambda \approx (C, D, \lambda)$,

$$E_\lambda(\mathcal{C}) = \lambda C(\mathcal{C}) + D(\mathcal{C}), \tag{6}$$

where $C(\mathcal{C}) = \sum_{n=1}^{N}\mathcal{H}(\mathcal{C}_n)$,

$$\begin{aligned} D(\mathcal{C}) \;=\; & \sum_{n=1}^{N}\int_{\Omega-\mathcal{C}_n}\Big(\tilde{f}(\vec{x},t_n) - f(\vec{x},t_n)\Big)^2 d\vec{x} \\ & + \sum_{n=1}^{N}\sum_{m=1}^{M}\int_{\Omega}\Big(f(\vec{x},t_n)\chi_{\mathcal{O}_{n,m}}(\vec{x}) \\ & - f(\phi_n(\vec{x}),t_{n+1})\chi_{\mathcal{O}_{n+1,m}}(\phi_n(\vec{x}))\Big)^2 d\vec{x}. \end{aligned}$$

According to remark 2.2, the minimizing of 2D+time affine energy (6) permit to compute a segmentation of video sequence, defined by a video partition, solution of the following optimization problem:

$$\min_P E_\lambda(\mathcal{C}), \tag{7}$$

where P is a video partition, and $\mathcal{C}$ is is the set of boundaries of the video partition P. In the following sections, we show how, we can compute a solution of this optimization problem on 2D+time hierarchy of video partitions.

3. OPTIMIZATION OF AN AFFINE ENERGY ON A HIERARCHY

Let $\varepsilon_\lambda \approx (C, D, \lambda)$ be an affine energy, defined on the set of partitions of domain X,

$$\begin{aligned} \varepsilon_\lambda : \quad Part(X) &\longrightarrow \mathbb{R}^+ \\ P &\longmapsto \lambda C(P) + D(P), \end{aligned} \tag{8}$$

where D and C are two functions on $Part(X)$, and $\lambda \in \mathbb{R}^+$. Find the partition $P \in Part(X)$ which minimizes the affine energy ε_λ is usually a difficult problem. However, if there exists a hierarchy of partitions of X, then the problem is easily solved by a dynamic programming algorithm [15].

DEFINITION 1 [16]. *Let $P \in Part(X)$. We say that H is hierarchy of partitions of X constructed over P if H is a family of nonempty subsets of X such that*

- *$X \in \mathcal{H}$;*
- *Any two sets in $\mathcal{H}$ are either nested or disjoint;*
- *Any set in H contains a set in P;*

We call X the root of hierarchy, the set $B(\mathcal{H}) = \{\{s\}\}_{s\in P}$ is called the base of the hierarchy. For any $x \in \mathcal{H}$, the subset $\mathcal{H}(x)$ of $\mathcal{H}$ determined by

$$\mathcal{H}(x) = \{y \in \mathcal{H} \quad | \quad y \subseteq x\}, \tag{9}$$

is also a hierarchy of partitions of x. We call $\mathcal{H}(x)$ the partial hierarchy on the subset x.

DEFINITION 2. *[15, 16] A cut of $\mathcal{H}$ is a partition of X whose elements are in $\mathcal{H}$.*

We note $\mathbb{C}ut(H)$ the set of cuts of $\mathcal{H}$. It is the set of partitions of X that we can build from $\mathcal{H}$. We shall assume that $\mathcal{H}$ has a finite number of elements. In this case, $\mathcal{H}$ is a tree whose nodes are the subsets of X in $\mathcal{H}$.

DEFINITION 3. *[16] We say that $G : Part(X) \to \mathbb{R}^+$ is separable, if there exists a function on the subsets of X which we denote by $\bar{G}$ such that:*

$$G(P) = \sum_{p\in P}\bar{G}(p), \qquad \forall P \in Part(X);$$

We say that G is subadditive if

$$\bar{G}(S \cup R) \leq \bar{G}(S) + \bar{G}(R), \qquad \forall S, R \in X, S \cap R = \emptyset;$$

We say that affine energy $\varepsilon_\lambda \approx (C, D, \lambda)$ is amultiscale energy, if C, D are separable and C is subadditive.

From now on we assume that ε_λ be an amultiscale energy. For any λ, let $\Gamma^*_\lambda(\mathcal{H})$ be the cut of $\mathcal{H}$ minimizing ε_λ,

$$\Gamma^*_\lambda(\mathcal{H}) = \arg\min_{\Gamma\in\mathbb{C}ut(\mathcal{H})}\varepsilon_\lambda(\Gamma). \tag{10}$$

Let $\mathcal{L}^*_\lambda(\mathcal{H})$ the set of nodes of $\mathcal{H}$ which are locally optimal in $\mathcal{H}$ for the energy ε_λ,

$$\mathcal{L}^*_\lambda(\mathcal{H}) = \{x \in \mathcal{H} \mid \forall y \in \mathbb{C}ut(H(x)),\ \varepsilon_\lambda(x) \leq \varepsilon_\lambda(y)\}. \tag{11}$$

For any $x \in \mathcal{H}$, let

$$\Delta^*(x) = \{\lambda |\ x \in \mathcal{L}^*_\lambda(\mathcal{H})\}. \tag{12}$$

The set $\Delta^*(x)$ represents the set of scales for which x is locally optimal in $\mathcal{H}$ for the amultiscale energy ε_λ. $\Delta^*(x)$ is an interval of the form $[a, +\infty)$ [15].

PROPOSITION 1. *[15, 16] For any $x \in \mathcal{H}$,*

- *$\Delta^*(x) = [\lambda^+(x), \lambda^-(x))$, where $\lambda^-(x) = \min_{s\in\mathcal{H}, x\subseteq s}\lambda^+(s)$.*

- $\Gamma^*_\lambda(\mathcal{H})) = \{x \in \mathcal{H}| \quad \lambda^+(x) \leq \lambda \leq \lambda^-(x)\}$.

We call the interval $[\lambda^+(x), \lambda^-(x))$ the interval of persistence of the region x, $\lambda^+(x)$ is the scale of apparition of node x, and $\lambda^-(x)$ is the scale of disappearance of node x.

DEFINITION 4. *[15] The persistent hierarchy obtained from $\mathcal{H}$ and ε_λ is*

$$\mathcal{H}^* = \{X \in \mathcal{H} : \Delta^*(X) \neq \emptyset\}. \tag{13}$$

REMARK 3.1. *[15] On the persistent hierarchy $\mathcal{H}^*$ we have*

$$\lambda^-(x) = \lambda^+(x^f), \tag{14}$$

where x^f denotes the father of x in $\mathcal{H}^*$.

Since ε_λ is an amultiscale energy, there exists two functions on the subsets of X which we denote by $\bar{G}_C$ and $\bar{G}_D$ such that:

- $C(P) = \sum_{p\in P} \bar{G}_C(p), \quad \forall P \in Part(X)$;
- $D(P) = \sum_{p\in P} \bar{G}_D(p), \quad \forall P \in Part(X)$;

For each node $x \in \mathcal{H}$, we define

$$\varepsilon_x : \lambda \mapsto \lambda \bar{G}_C(x) + \bar{G}_D(x). \tag{15}$$

We define also the partial energy $\varepsilon^*_x(\lambda)$ of the node $x \in \mathcal{H}$ as the energy of the optimal cut of the partial hierarchy $\mathcal{H}(x)$:

$$\varepsilon^*_x(\lambda) = \varepsilon_\lambda(\Gamma^*_\lambda(\mathcal{H}(x))). \tag{16}$$

Observe that for any element of base $B(\mathcal{H}) = \{\{s\}\}_{s\in X}$ of the hierarchy, we have

$$\varepsilon^*_s(\lambda) = \varepsilon_s(\lambda). \tag{17}$$

PROPOSITION 2. *[15, 16] The partial energies $\varepsilon^*_x(\lambda)$ of the nodes of $\mathcal{H}$ are related by the dynamic programming equation:*

$$\varepsilon^*_x(\lambda) = \inf\{\varepsilon_x(\lambda); \sum_{s\in\mathcal{F}(x)} \varepsilon^*_s(\lambda)\}, \tag{18}$$

where $\mathcal{F}(x)$ is the family of children of x.

PROPOSITION 3. *[15, 16] For any $x \in \mathcal{H}$,*

$$\varepsilon^*_x(\lambda) = \begin{cases} \sum_{s\in\mathcal{F}(x)} \varepsilon^*_s(\lambda), & \lambda < \lambda^+(x); \\ \varepsilon_x(\lambda), & \lambda \geq \lambda^+(x); \end{cases} \tag{19}$$

where $\mathcal{F}(x)$ is the family of children of x.
If C is strictly subadditive, then $\lambda^+(x) \in \mathbb{R}$ and is the only solution of

$$\varepsilon_x(\lambda) = \sum_{s\in\mathcal{F}(x)} \varepsilon^*_s(\lambda). \tag{20}$$

4. VIDEO SEGMENTATION ALGORITHM

In this section we propose a new segmentation method of video sequence. The idea is the computing a video segmentation by selecting a 2D+time partition of $\Omega \times [T_i, T_f]$, using the minimization of the affine energy E_λ on a 2D+time persistent hierarchy. In the first we give a sufficient condition to guarantee that the 2D+time affine energy associated to our extended Mumford-Shah Functional $E_\lambda(\mathcal{C}) = \lambda C(\mathcal{C}) + D(\mathcal{C})$, is amultiscale energy.

PROPOSITION 4. *If the measure $\mathcal{H}$ is separable and strictly subadditive, then the 2D+time affine energy $E_\lambda(\mathcal{C}) = \lambda C(\mathcal{C}) + D(\mathcal{C})$ is amultiscale energy, and C is strictly subadditive.*

Proof
Let us prove first that, the function C is separable and strictly subadditive. Since $\mathcal{H}$ is strictly subadditive , there exists a function $\bar{G}$ of the subsets of Ω, such that, for any set of boundaries β of a partition P of image domain Ω:

$$\mathcal{H}(\beta) = \sum_{p\in P} \bar{G}(p) \tag{21}$$

$$\bar{G}(p \cup p') < \bar{G}(p) + \bar{G}(p') \quad \forall p, p' \in Part(\Omega), p \cap p' = \emptyset. \tag{22}$$

Let $\mathcal{C} = \{\mathcal{C}_n\}_{n\in[1,N]}$ be the set of boundaries of 2D+time partition $V = \{v_m\}_{m\in[1,M]}$ of $\Omega \times [T_i, T_f]$. The 2D+time section v_m is defined by 2D sections $\mathcal{O}_{n,m} \subseteq \Omega$ at each time t_n,

$$v_m = \bigcup_{n=1}^{N} \mathcal{O}_{n,m}. \tag{23}$$

Hence, in each time t_n the set $\mathcal{C}_n$ represents the boundaries of the partition $\{\mathcal{O}_{n,m}\}_{m\in[1,M]}$ of image domain Ω. According to the property (21), we verify that

$$C(\mathcal{C}) = \sum_{n=1}^{N} \mathcal{H}(\mathcal{C}_n) = \sum_{n=1}^{N}\sum_{m=1}^{M} \bar{G}(\mathcal{O}_{n,m}) = \sum_{m=1}^{M} \bar{\bar{G}}(v_m), \tag{24}$$

where $\bar{\bar{G}}$ is function of the subsets of $\Omega \times [T_i, T_f]$, defined by:

$$\bar{\bar{G}}(v_m) = \sum_{n=1}^{N} (\bar{G}(\mathcal{O}_{n,m}). \tag{25}$$

Let $v_m, v_{m'}$ be two subset of $\Omega \times [T_i, T_f]$. According to the property (22), we verify that

$$\bar{\bar{G}}(v_m \cup v_{m'}) = \sum_{n=1}^{N} \bar{G}(\mathcal{O}_{n,m} \cup \mathcal{O}_{n,m'}) < \bar{\bar{G}}(v_m) + \bar{\bar{G}}(v_{m'}). \tag{26}$$

Then the function C is separable and strictly subadditive.
Now, prove that the function D is separable. Let $\bar{R}$ be a function of the subsets of $\Omega \times [T_i, T_f]$, defined by

$$\bar{R}(v_m) = \sum_{n=1}^{N} \int_{\mathcal{O}_{n,m}} \Big(\tilde{f}(\vec{x}, t_n) - f(\vec{x}, t_n)\Big)^2 d\vec{x} + \sum_{n=1}^{N} \int_{\Omega} \Big(f(\vec{x}, t_n)\chi_{\mathcal{O}_{n,m}}(\vec{x}) - f(\phi_n(\vec{x}), t_{n+1})\chi_{\mathcal{O}_{n+1,m}}(\phi_n(\vec{x}))\Big)^2 d\vec{x}.$$

Hence $D(\mathcal{C}) = \sum_{m=1}^{M} \bar{R}(v_m)$.
Then the 2D+time affine energy E_λ is amultiscale energy. □
Therefore, if the measure $\mathcal{H}$ is separable and strictly subadditive, we can use the theory of optimization of an affine energy on a hierarchy, for gave a solution of optimization problem (7).
let $\bar{\mathcal{V}} = \{\bar{V}_m\}_{m\in[1,M]}$ be a 2D+time persistent hierarchy of video domain $\Omega \times [T_i, T_f]$. In the following,, we assume that the measure $\mathcal{H}$ is separable and strictly subadditive, then $E_\lambda(\mathcal{C}) = \lambda C(\mathcal{C}) + D(\mathcal{C})$ is amultiscale energy, and C is strictly subadditive, and the combining of the results of Propositions 1, 2 and 3 permit to construct an algorithm of video segmentation, that can compute a solution for the optimization problem (7) on $\bar{\mathcal{V}}$. The implementation of this video segmentation algorithm is based on four steps. In the first step we compute the partial energy E^*_m by a rise on the tree of $\bar{\mathcal{V}}$, using the dynamic programming equation of Proposition 2. We initialize the computing with the partial energy of base elements of $\bar{\mathcal{V}}$, given by equation (17). In second step we compute the scale of apparition for each node. Since C is strictly subadditive, the scale of apparition for each node are calculated by intersection between E_m and $\sum_{s\in\mathcal{F}(m)} E^*_s$ (Proposition 3). In third step we compute the scale of disappearance for each node by a down on the tree of $\bar{\mathcal{V}}$, using Proposition 1.

In fourth step we store the nodes which satisfy the condition of Proposition 1. Finally we compute the optimal cut $\Gamma^*_\lambda(\vec{\mathcal{P}})$, that is a solution for the optimization problem of video segmentation (7).

Video segmentation algorithm

INPUT: $\vec{\mathcal{P}} = \{\bar{V}_m\}_{m\in[1,M]}$: 2D+time persistent hierarchy.
Step 1:
for $m = 1$ to M
do { compute E^*_m using the dynamic programming equation of Proposition 2}
end for
Step 2:
for $m = 1$ to M
do { compute λ^+_m using the intersection between E_m and $\sum_{s\in\mathcal{F}(m)} E^*_s$ (Proposition 3)}
end for
Step 3:
for $m = 1$ to M
do { comput λ^-_m using Proposition 1}
end for
Step 4:
for $m = 1$ to M
if $\{\lambda^+_m \leq \lambda \leq \lambda^-_m\}$
do { Store $\bar{V}_m$ }
end if
end for
OUTPUT: $\Gamma^*_\lambda(\vec{\mathcal{P}}) = \{\bar{V}Stored\}$ (Proposition 1).

Therefore, if we can build a 2D+time persistent hierarchy, we can use this video segmentation algorithm for compute a 2D+time segmentation of video sequence. In the following sections, we show, how we can build a 2D+time persistent hierarchy of video domain.

5. 2D+TIME PERSISTENT HIERARCHY.

Now we propose an approach to build a 2D+time persistent hierarchy of $\Omega \times [T_i, T_f]$. Our approach is based on the 2D-shapes of frames of video sequence, and Scale-invariant feature transform (SIFT).

5.1 Tree of image shapes

The shapes of an image are built from the connected components of level sets. It is well known that connected components of level sets have a tree structure [16][17, 18]. Image f is characterized by its upper (lower) level sets

$$[f \geq \mu] = \{p \in \Omega, f(p) \geq \mu\}; \tag{27}$$

$$[f < \mu] = \{p \in \Omega, f(p) < \mu\}. \tag{28}$$

DEFINITION 5. *[16] We say that* ***sat*** $: Part(\Omega) \rightarrow Part(\Omega)$ *is operator of saturation if:*

- $\forall A \subset \Omega$, $\Omega \backslash$***sat***(A) *is either* $\emptyset$ *; or a connected component of* $\Omega \backslash A$*;*
- $\forall A \subset \Omega$, ***sat***$(\Omega \backslash$***sat***$(A)) = \Omega$ *or* $\emptyset$*;*
- $A \subset B \Longrightarrow$ ***sat***$(A) \subset$ ***sat***(B);
- ***sat*** $\circ$ ***sat*** $=$ ***sat***.

DEFINITION 6. *We call shapes of inferior (resp. superior) type the sets*

$$\textbf{\textit{sat}}(CC([f < \mu])) \tag{29}$$

$$\textbf{\textit{sat}}(CC([f \geq \mu]))). \tag{30}$$

THEOREM 7. *[16] Any two shapes are either disjoint or nested.*

From this result, we can conclude that the set of shapes of an image has an inclusion tree structure. For simplicity, we assume that our image is discrete. Then we can represent the tree as a finite structure; the shapes are the tree nodes and the parent-child relation- ship, represented by the links between nodes, is determined by inclusion (the child A being a shape contained in the father A^f with no other shape B such that $A \in B \in A^f$). The root of the tree is

$$\bar{\Omega} = Sat([f \geq min_\Omega f]) \tag{31}$$

5.2 Scale-invariant feature transform (SIFT).

The Scale-invariant feature transform is an algorithm in computer vision to detect and describe local features in images. The algorithm was published by David Lowe in 1999 [19]. this algorithm transforms image data into scale-invariant coordinates relative to local features. An important aspect of this transform is that it generates large numbers of features that densely cover the image over the full range of scales and locations [20]. Following are the major stages of Scale-invariant feature transform [20]:

• **Scale-space extrema detection:** The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points, which are called keypoints in the SIFT framework.

• **Keypoint localization:** Scale-space extrema detection produces too many keypoint candidates. At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability.

• **Orientation assignment:** In this step, each keypoint is assigned one or more orientations based on local image gradient directions. This is the key step in achieving invariance to rotation as the keypoint descriptor can be represented relative to this orientation and therefore achieve invariance to image rotation.

• **Keypoint descriptor:** The previous operations have assigned an image location, scale, and orientation to each keypoint. These parameters impose a repeatable local 2D coordinate system in which to describe the local image region, and therefore provide invariance to these parameters. The next step is to compute a descriptor for the local image region that is highly distinctive yet is as invariant as possible to remaining variations, such as change in illumination or 3D viewpoint.

• **Keypoints matching:** Given a set of keypoint descriptors computed from two different images, these keypoint descriptors can be mutually matched by for each point finding the point in the other image domain that minimizes the Euclidean distance between the descriptors represented as 128-dimensional vectors. To suppress matches that could be regarded as possibly ambiguous, Low only accepted matches for which the ratio between the distances to the nearest and the next nearest points is less than 0.8. Fig 1 gives the result of keypoints matching obtained by by Lowes software [20].

5.3 2D hierarchy

5.3.1 ***Merging algorithm*** . Let $\mathcal{S}(f_n)$ be the tree of the shapes of image $f_n = f(., t_n)$ of video sequence $\{f(\vec{x}, t_n)\}_{1\leq n\geq N}$, and $\mathcal{P}_n$ be a partition of f_n in $\mathcal{S}(f_n)$. We suppose here that the tree of the shapes have a finite number of elements, then $\mathcal{P}_n = \bigcup_{j=1}^{J} R_{n,j}$, where $R_{n,j}$ are the regions of the partition $\mathcal{P}_n$. For any region $R_{n,j}$, let $\{B_1, ..., B_p\}$ be a set of sibling regions of $R_{n,j}$, and A be the father of $R_{n,j}$ in the tree of the shapes $\mathcal{S}(f_n)$. We define a new partition $\mathcal{P}'_n$ by merging the re-

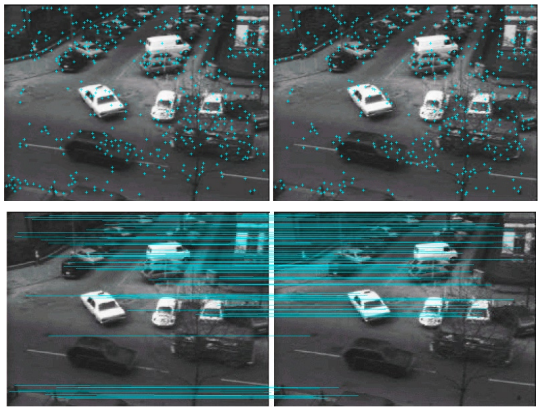

Fig. 1. The top row presents keypoints of two images of video sequence. The keypoints matching are show in last row.

gions $\{B_1, ..., B_p, R_{n,j}\}$.

$$\mathcal{P}'_n = \Big\{\mathcal{P}_n \backslash \{B_1, ..., B_p, R_{n,j}\}\Big\} \cup A \tag{32}$$

let $\Delta\bar{E}(\mathcal{P}_n, R_{n,j}) = \bar{E}_\lambda(\mathcal{P}_n) - \bar{E}_\lambda(\mathcal{P}'_n)$.
Where

$$\bar{E}_\lambda(\mathcal{P}_n) = \lambda\mathcal{H}(\mathcal{P}_n) + \int_{\mathcal{P}_n} \Big(\tilde{f}(\vec{x}, t_n) - f(\vec{x}, t_n)\Big)^2 d\vec{x} \tag{33}$$

merging algorithm

INPUT: $\mathcal{P}_{0,n} = \bigcup_{j=1}^{J} R_{n,j}$.
Step 1:
for $j = 1$ to J
do { compute $\Delta\bar{E}(\mathcal{P}_{0,n}, R_{n,j})$ and insert $R_{n,j}$ in a queue $\mathcal{Q} = \{q_j, j = 1, ..., J\}$ with priority $\Delta\bar{E}(\mathcal{P}_{0,n}, R_{n,j})$
end for
Step 2: Iterate the following procedure
if $\Delta E(\mathcal{P}_{i,n}, q_1) > 0$
do { define $\mathcal{P}_{i+1,n} = \{\mathcal{P}_{i,n} \backslash \{B_1, ..., B_p, q_1\}\} \cup A$
end if
We stop when no node q_1 exists with $\Delta\bar{E}(\mathcal{P}_{i,n}, q_1) > 0$.
OUTPUT: $\bar{\mathcal{P}}_n = \mathcal{P}_{i_f,n}$ is a partition.

The last partition obtained $\bar{\mathcal{P}}_n$ determines by the father regions of the regions $\{R_{n,j}\}_{j\in[1,J]}$.

5.3.2 ***Building of 2D hierarchy*** . For each image f_n, we start with initial scale value λ_0, and a initial partition $\bar{\mathcal{P}}_{0,n} = \mathcal{P}_{0,n}$ of $\mathcal{S}(f_n)$. Let $\bar{\mathcal{P}}_{1,n}$ be the partition obtained by merging algorithm in $\bar{\mathcal{P}}_{0,n}$. We continue iteratively the process of merging algorithm by minimizing the function $\bar{E}_{\lambda_{k+1}}$, $\lambda_{k+1} = 2\lambda_k$ on the partition $\bar{\mathcal{P}}_{k,n}$. The iterative process may be stopped either when the value of λ_k attains a maximum scale value λ_{K-1}, finally we add the root partition $\bar{\mathcal{P}}_{K,n} = \Omega$.
Since, the partitions obtained $\{\bar{\mathcal{P}}_{k,n}\}_{0\le k\le K}$ are defened by shapes of the tree $\mathcal{S}(f_n)$, and any two shapes are either disjoint or nested, the partitions obtained $\{\bar{\mathcal{P}}_{k,n}\}_{0\le k\le K}$ determine a 2D hierarchy of image f_n.

5.4 2D+time hierarchy

Now we show, how we can build a 2D+time hierarchy of video domain. In the first step, for the frames f_1 and f_2 we built a 2D hierarchys defined by K partitions $\{\bar{\mathcal{P}}_{k,1}\}_{0\le k\le K}$ and $\{\bar{\mathcal{P}}_{k,2}\}_{0\le k\le K}$ obtained by merging algorithm, and we compute the keypoints $\{p_{1,i}\}_{1\le i\le I_1}$ and $\{p_{2,i}\}_{1\le i\le I_2}$ in the SIFT framework of image f_1 and f_2, and for all keypoint $\{p_{1,i}\}_{1\le i\le I_1}$ we search the matching keypoints in $\{p_{2,i}\}_{1\le i\le I_2}$. The matching between two keypoints of two images is found with minimum euclidean distance for the invariant descriptor vector.
In the second step, we compute the temporarily connected regions in $\{\bar{\mathcal{P}}_{k,1}\}_{0\le k\le K}$ and $\{\bar{\mathcal{P}}_{k,2}\}_{0\le k\le K}$:

- We compute the similar regions in $\bar{\mathcal{P}}_{0,1}$ and $\bar{\mathcal{P}}_{0,2}$ and we temporarily connect these similar regions.
- Each non-similar region $R_{0,1,j} \in \bar{\mathcal{P}}_{0,1}$ is characterized by a set $\{p_{1,i}\}_{1\le i\le I_{0,j}}$ of keypoints $\{p_{1,i}\}_{1\le i\le I_1}$. Let $\{p'_{2,i}\}_{1\le i\le I'_{0,j}}$ be the matching keypoints of keypoints $\{p_{1,i}\}_{1\le i\le I_{0,j}}$ in image f_2. Let $\mathcal{M}(R_{0,1,j})$ be the set of matching regions of $R_{0,1,j}$ in $\bar{\mathcal{P}}_{0,2}$,

$$\mathcal{M}(R_{0,1,j}) = \Big\{R_2 \in \bar{\mathcal{P}}_{0,2} : \exists p \in \{p'_{2,i}\}_{1\le i\le I'_{0,j}}, p \in R_2\Big\}. \tag{34}$$

We note $\mathfrak{F}(\mathcal{M}(R_{0,1,j}))$ the father regions of the set regions $\mathcal{M}(R_{0,1,j})$ in the 2D hierarchy $\{\bar{\mathcal{P}}_{k,2}\}_{0\le k\le K}$. and we define a new partition $\bar{\mathcal{P}}'_{0,2}$ by replacing the set of matching regions $\mathcal{M}(R_{0,1,j})$ by $\mathfrak{F}(\mathcal{M}(R_{0,1,j}))$ and we temporarily connect connect $R_{0,1,j}$ with $\mathfrak{F}(\mathcal{M}(R_{0,1,j}))$.

We continue this process for computing the temporarily connected partitions $\bar{\mathcal{P}}'_{k,2}$ for each partition level $k \in \{1, ..., K\}$.
We continue the process of the first and second step for computing the temporarily connected partitions between each two successive frames f_n and f_{n+1}. We observe that, For each time t_n, The last partitions obtained $\{\bar{\mathcal{P}}'_{k,n}\}_{1\le k\le K}$ determine a 2D hierarchy of Ω, and for each partition level $k \in \{1, ..., K\}$, the regions of $\{\bar{\mathcal{P}}'_{k,n}\}_{1\le n\le N}$ are temporarily connected. So $\mathcal{V}_0 = \{\bar{\mathcal{P}}'_{k,n}\}_{1\le k\le K, 1\le n\le N}$ is a 2D+time hierarchy of $\Omega \times [T_i, T_f]$.

5.5 Transform the 2D+time hierarchy to 2D+time persistent hierarchy

let $\mathcal{V}_0 = \{v_s\}_{1\le s\le S}$ be the 2D+time hierarchy of $\Omega \times [T_i, T_f]$ given in precedent section. The nodes non-persistence of $\mathcal{V}_0$ are not important for the minimization of amultiscale energy E_λ on 2D+time hierarchy $\mathcal{V}_0$. So it is natural to remove these nodes and transform 2D+time hierarchy $\mathcal{V}_0$ to a 2D+time persistent hierarchy by the following Greedy algorithm.
let $\Delta E(\mathcal{V}, v_s) = E_\lambda(\mathcal{V}) - E_\lambda(\mathcal{V} \setminus \{v_s\})$.

Greedy algorithm

INPUT: $\mathcal{V}_0 = \{v_s\}_{j\in[1,S]}$.
Step 1:
for $s = 1$ to S
do { compute $\Delta E(\mathcal{V}_0, v_s)$ and insert v_s in a queue $\mathcal{Q} = \{q_n, n = 1, ..., S\}$ with priority $\Delta E(\mathcal{V}_0, v_s)$}
end for
Step 2: Iterate the following procedure
if $\Delta E(\mathcal{V}_i, q_1) > 0$
do { define $\mathcal{V}_{i+1} = \mathcal{V}_i \setminus q_1$ }
end if
We stop when no node q_1 exists with $\Delta E(\mathcal{V}_i, q_1) > 0$.
OUTPUT: $\mathcal{V}_i$ is a persistent hierarchy.

The last tree obtained $\mathcal{V}_i$ determines a 2D+time persistent hierarchy. It is a local optimal solution of 7, in the sense that any other merging of regions increases the energy.

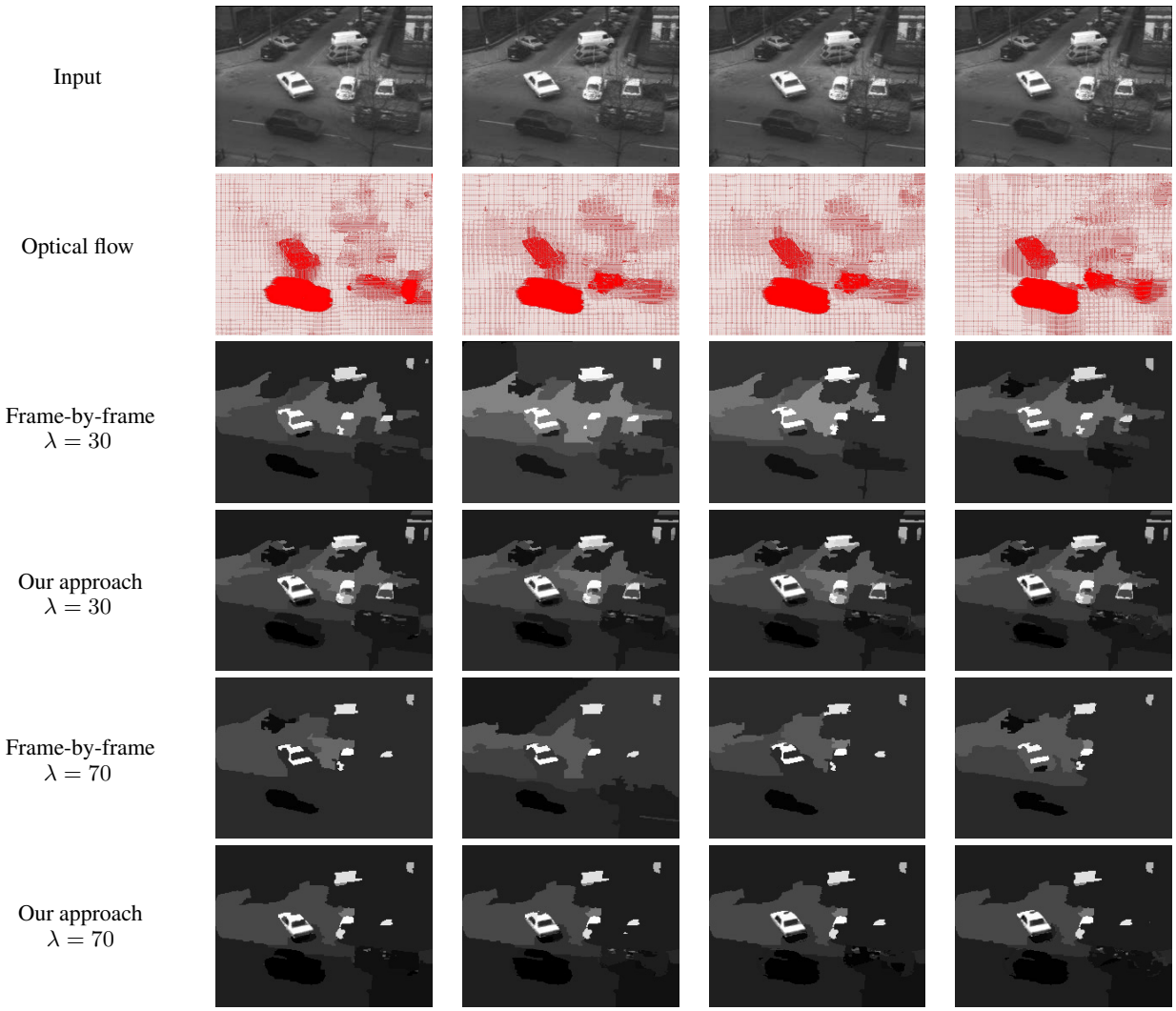

Fig. 2. Video segmentation results of Hamburg taxi sequence with various values of λ. The top row presents four frames of video sequence, the motion fields (optical flow) of these four frames are showed in second row. The remaining four rows illustrate 2D Mumford-Shah frame-by-frame segmentation and result of the proposed approach, with two values of λ.

6. EXPERIMENTAL RESULTS

In this section we present some experimental results obtained using the video segmentation procedure described. In the first we compared the results derived by our method with those obtained by 2D Mumford-Shah frame-by-frame approach. Fig 2 demonstrates segmentation results of Hamburg taxi sequence with various values of λ. The top row presents four frames of video sequence, the motion fields (optical flow) of these four frames are showed in second row. The third row illustrates 2D Mumford-Shah frame-by-frame segmentation, the result of the proposed approach is show in last row. We observe that in our video segmentation, it is possible to follow the movements. Unlike, the 2D Mumford-Shah frame-by-frame approach give a segmentation without considering the movement, the reason is the absence of the movements in 2D Mumford-Shah Functional. and we observe that With our approach, modes span space and time and temporal coherence is naturally achieved because our video segmentation is obtained by selecting a Spatio-temporal partition on 2D+time hierarchy that ensures temporal and spatial connectedness of regions.
In fig 3 and 4, we compare our results of Monkey bar and Dance sequences, against others on the leading methods that treat the video as a 3D space-time volume; Streaming Mean-Shift approach (SMS) [22], and Hierarchical Graph-Based (HGB) [21]. Fig 3 tests on fast moving footage containing small objects. We observe that the fine scale features are retained when they are in motion (man's face) Unlike to MSN and HGB methods these fine scale features are absent, and our segmentation successfully separates the motion layers while providing more details than other approaches (compare footwear and hat of the man). Similarly, in fig 4 our segmentation successfully separates the motion layers while providing more details than other approaches, and we observe that in our segmentation method spatio-temporal coherence is naturally achieved, this is shown by temporal stability of the background.

7. CONCLUSION

We have developed a 2D+time extension of Mumford-Shah Functional, and a new approach of construction of a 2D+time hierarchy, this approach is based on the 2D-shapes of images of video sequence, and Scale-invariant feature transform (or SIFT). The minimization of 2D+time extension of Mumford-Shah Functional on this hierarchy permit the computing a video segmentation by selecting a partition of video domain.

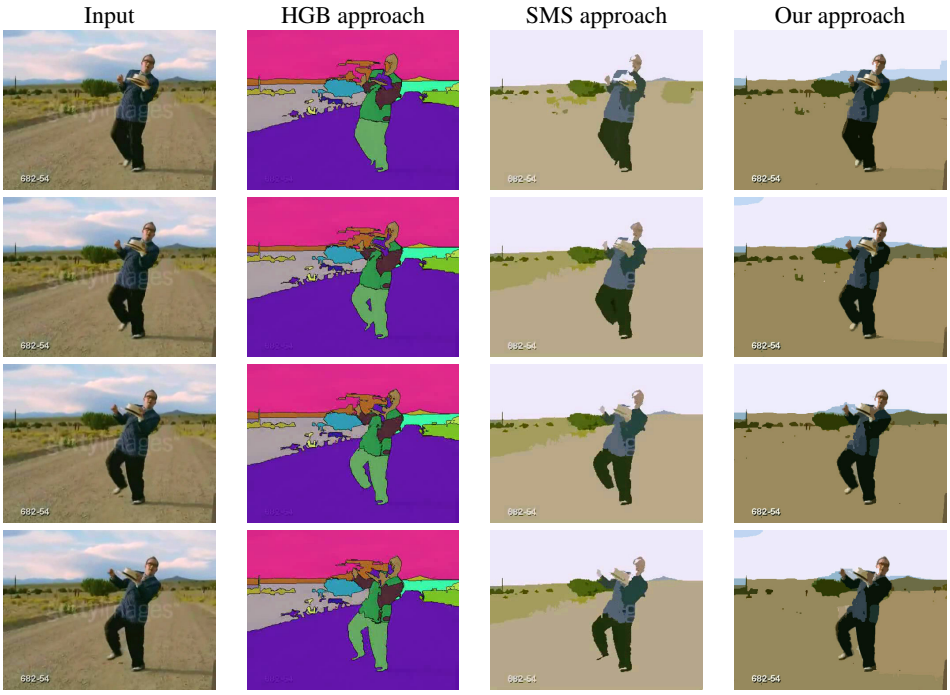

Fig. 3. Comparing the accuracy and coherence of the proposed approach on the Dance sequence, to Hierarchical Graph-Based approach (HGB) [21], and Streaming Mean-shift Approach (SMS) [22].

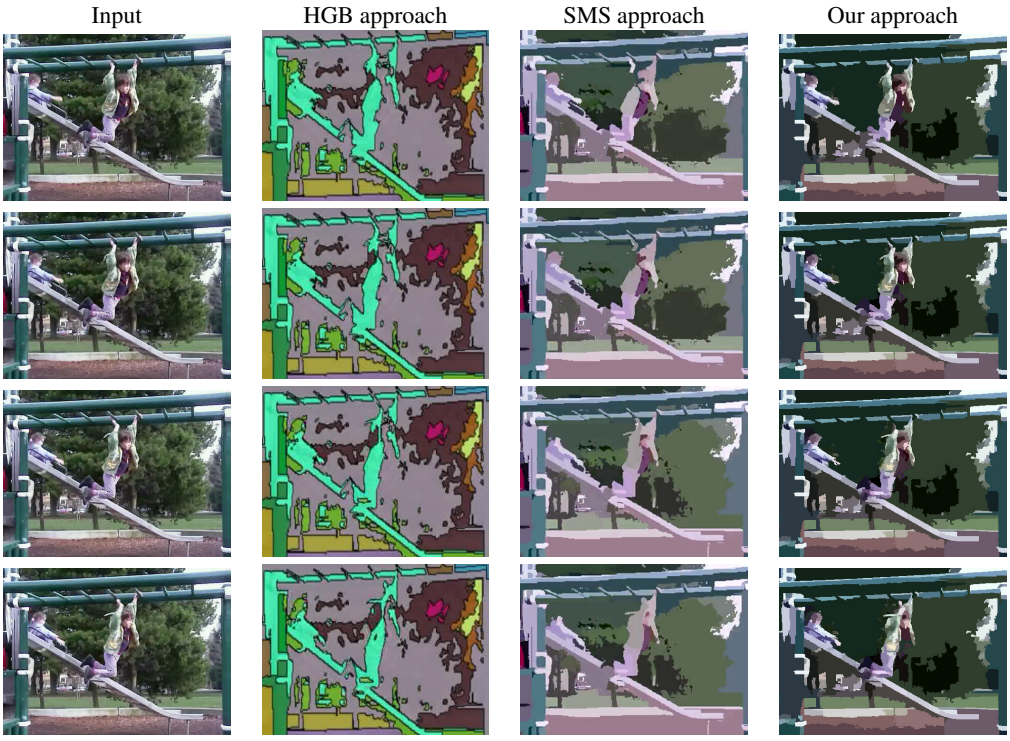

Fig. 4. Comparing the accuracy and coherence of the proposed approach on the Monkey bar sequence, to Hierarchical Graph-Based approach (HGB) [21], and Streaming Mean-shift Approach (SMS) [22].

8. REFERENCES

[1] H. Winnemoller, S.C. Olsen, and B. Gooch. Real-time video abstraction. *ACM Transactions on Graphics Proc. of the ACM SIGGRAPH conf*, 25:1221–1226, July 2006.

[2] J. Chen, S. Paris, and F. Durand. Real-time edge-aware image processing with the bilateral grid. *ACM Transactions on Graphics Proc. of the ACM SIGGRAPH conf*, 26:103, July 2007.

[3] Y. Wang, K.F Loe, T. Tan, and J-K. Wu. Spatiotempo-

ral video segmentation based on graphical models. *IEEE transactions on image processing*, 14:937–947, July 2005.

[4] J.P Collomosse, D. Rowntree, and P.M Hall. Stroke surfaces: Temporally coherent artistic animations from video. *IEEE Transactions on Visualization and Computer Graphics*, 11:540–549, 2005.

[5] W. Brendel and S. Todorovic. Video object segmentation by tracking regions. *IEEE 12th International Conference on Computer Vision*, pages 833–840, 2009.

[6] D. Comaniciu and P. Meer. A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 24:603–619, 2002.

[7] A. W. Klein, P. J. Sloan, A. Finkelstein, and M. F. Cohen. Stylized video cubes. pages 15–22.

[8] J. Wang, B. Thiesson, Y. Xu, and M. Cohen. Image and video segmentation by anisotropic kernel mean shift. *in Proc. 8th European Conference on Computer Vision, Prague, Czech Republic*, 2:238–249, May 2004.

[9] D. DeMenthon and R. Megret. Spatio-temporal segmentation of video by hierarchical mean shift analysis. *Technical Report: LAMP-TR-090/CAR-TR-978/CS-TR-4388/UMIACS-TR-2002-68, University of Maryland, College Park*, 2002.

[10] S. Paris and F. Durand. A topological approach to hierarchical segmentation using mean shift. *IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, Minnesota, USA*, pages 1–8, June 2007.

[11] T. Wang, J.-Y. Guillemaut, and J. Collomosse. Multi-label propagation for coherent video segmentation and artistic stylization. *17th IEEE International Conference on Image Processing (ICIP)*, pages 3005–3008, September 2010.

[12] H. Greenspan, J. Goldberger, and A. Mayer. A probabilistic framework for spatio-temporal video representation. *CCV '02 Proceedings of the 7th European Conference on Computer Vision-Part IV*, pages 461–475, 2002.

[13] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–686, 1989.

[14] L. Vese and T. Chan. A mumtiphase level set framework for image segmentation using the mumford and shah model. *Inter. J. Computer Vision*, 50:271–293, 2002.

[15] L. Guigues. Modles multi-chelles pour la segmentation d'images. *PhD thesis, Cergy-Pontoise University*, 2003.

[16] L.I. Muoz. Image segmentation and compression using the tree of shapes of an image. motion estimation. *PhD thesis, Pompeu Fabra University, Barcelona*, 2005.

[17] C. Ballester V. Caselles and P. Monasse. The tree of shapes of an image. *SAIM: Control, Optimization and Calculus of Variations*, 9:1–18, 2003.

[18] P. Monasse. Morphological representation of digital images and application to registration. *PhD thesis, Universit Paris IX-Dauphine*, June 2000.

[19] D. G. LOWE. Object recognition from local scale-invariant features. *ICCV '99 Proceedings of the International Conference on Computer Vision*, 2:1150–1157, 1999.

[20] D. G. LOWE. Distinctive image features from scale-invariant keypoints. *International Journal of Computer vision*, 60:91–110, November 2004.

[21] M. Grundmann, V. Kwatra, M. Han, and I. Essa. Efficient hierarchical graph-based video segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, USA*, June 2010.

[22] S. Paris. Edge-preserving smoothing and mean-shift segmentation of video streams. *ECCV '08 Proceedings of the 10th European Conference on Computer Vision: Part II*, pages 460–473, 2008.