Spatio-temporal Segmentation with Mumford-Shah Functional

Mohamed El Aallaoui
Laboratory of Mathematical Engineering (LINMA),
Department of Mathematics and Computer Science,
Faculty of Sciences -Eljadida-Morocco
Email: m_elallaoui@yahoo.fr

Abdelvahad Gourch
Faculté des Sciences Juridiques, Économiques
et Sociales de Ain Sebaâ
-Casablanca-Morocco
Email: agourch2002@yahoo.fr

Abstract—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 have extended the image segmentation methods to 3D domain. In these volumetric approaches, it is not known a priori, which regions to track, what frames contain those regions, or the time-direction for tracking (forward or backward). In this paper we present an efficient and scalable method for spatio-temporal segmentation obtained by minimizing a 2D+time extension of the simplified Mumford-Shah functional.

I. EXTENDING MUMFORD-SHAH FUNCTIONAL FOR 2D+TIME SEGMENTATION

The Mumford Shahan functional was introduced by Mumford and Shah in 1989 [1]. It follows:

\[ MS_{\lambda}(\hat{f}, C) = \lambda \mathcal{H}(C) + \int_{\Omega - C} (\hat{f}(x) - f(x))^2 dx. \]

Like before, \( f \) is our image function. We have \( \Omega = \Omega_1 \cup \Omega_2 \cup ... \cup \Omega_n \cup C \) in which \( \Omega \) is the domain of our image, \( \Omega_1 \) is the region in our image that represents a section \( C \), which does not include the boundaries, and \( C \) is the set of smooth arcs that make up boundaries for the \( \Omega \). \( \mathcal{H}(C) \) denotes the length of the system of curves \( C \). The function \( \hat{f} \) is a piecewise constant image, and \( \lambda > 0 \) is a parameter. Mumford-Shah segmentation of image \( f \) is defined by a pair \((C, \hat{f})\), which minimize the Mumford Shahan functional \( MS_{\lambda}(\hat{f}, C) \).

Let \( \tilde{f} : \Omega \times [T_i, T_f] \rightarrow \mathbb{R} \) be a video sequence with spatial domain \( \Omega \) 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((\tilde{x}, t_n) \] defined by a pair \((C, \tilde{f})\), such that:

- \( \tilde{f} \) is piecewise regular function in \((\Omega \times [T_i, T_f]) \setminus C\):
  \[ \tilde{f} : (\Omega \times [T_i, T_f]) \setminus C \rightarrow \mathbb{R} \]

  \( \tilde{f} \) is a 2D+time extension of the simplified Mumford-Shah functional obtainned by minimizing the 2D+time extension of the simplified Mumford-Shah functional as an anmscalar energy, which is minimized on a hierarchy of video domain. The construction of this hierarchy based on the 2D-shapes of video images, and Scale-invariant feature transform (SIFT).

\[ E_{\lambda}(\tilde{f}, C) = \lambda \sum_{n=1}^{N} \mathcal{H}(C_n) + \sum_{n=1}^{N} \int_{\bar{C}_n} (\tilde{f}(\bar{x}, t_n) - f(\bar{x}, t_n))^2 d\bar{x} + \]

\[ \sum_{n=1}^{N} \sum_{m=1}^{M} \int_{\Omega} \chi_{\Omega_{n,m}}(\bar{x}) f(\bar{x}, t_n) \chi_{\Omega_{n+1,m}}(\phi_{n}(\bar{x}))^2 d\bar{x}, \]

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

- \( C \) is the set of boundaries where \( \tilde{f} \) is discontinuous.

Notice that, the set of boundaries \( C \) 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 \( O_{n,m} \) at each time \( t_n \).

\[ v_m = \bigcup_{n=1}^{N} O_{n,m}. \]

Remark 1.1: At each time \( t_n \), the set \( \{O_{n,m}\}_{m\in[1,M]} \) is a partition of \( \Omega \), that generates a set of boundaries. We note \( C_n \) this set of boundaries.

Hence, the \( C \) is defined by the sets \( C_n \).

\[ C = \bigcup_{n=1}^{N} C_n. \]

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

\[ E_{\lambda}(\tilde{f}, C) = \lambda \sum_{n=1}^{N} \mathcal{H}(C_n) + \sum_{n=1}^{N} \int_{\Omega - C_n} (\tilde{f}(\bar{x}, t_n) - f(\bar{x}, t_n))^2 d\bar{x} + \]

\[ \sum_{n=1}^{N} \sum_{m=1}^{M} \int_{\Omega} \chi_{\Omega_{n,m}}(\bar{x}) f(\bar{x}, t_n) \chi_{\Omega_{n+1,m}}(\phi_{n}(\bar{x}))^2 d\bar{x}, \]

where, \( \lambda \in \mathbb{R}^+ \), \( \mathcal{H}(C_n) \) is measure of \( C_n \) and \( \phi_{n}(\bar{x}) \) represents the trajectory of the particle which was in the position \( \bar{x} \) at time \( t_n \), and corresponds \( t_{n+1} \). We modelize the trajectory \( \phi_{n} \) with an affine model defined by:

\[ \phi_{n}(x_1, x_2) = (u_n+\delta_n t + x_1, v_n+\delta_n t + x_2), \]

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 1.2: We observe that, given \( C \), the minimum of \( E_{\lambda}(\tilde{f}, C) \) with respect to the variable \( \tilde{f}_m \) is explicitly given by:

\[ \tilde{f}_m = \frac{1}{\sum_{n=1}^{N} |O_{n,m}|} \sum_{n=1}^{N} \int_{O_{n,m}} f(\bar{x}, t_n) d\bar{x}. \]
This observation permits to write our model of 2D+time Mumford-Shah functional $E_h(f, C)$, as a 2D+time affine energy of $C$ and denote it by $E_h \approx (C, D, \lambda)$,

$$E_h(C) = \lambda C(C) + D(C), \quad (7)$$

where

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

$$D(C) = \sum_{n=1}^{N} \int_{\Omega - C_n} (\bar{f}(x, t_n) - f(x, t_n))^2 \, dx + \sum_{n=1}^{N} \sum_{m=1}^{M} \int_{\Omega} (f(x, t_n) \chi_{O_{n,m}}(x) - f(\phi_n(x), t_{n+1}) \chi_{O_{n+1,m}}(\phi_n(x))) \, dx,$$

According to remark 1.2, the minimizing of $2D+time$ affine optimization problem: $\min \lambda \in \mathbb{R}^+$ according to remark 1.2, the minimizing of $2D+time$ affine energy $E_h(C)$ permit to compute a segmentation of video sequence, defined by a video partition, solution of the following optimization problem:

$$\min_P E_h(C), \quad (8)$$

where $P$ is a video partition, and $C$ 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.

II. OPTIMIZATION OF AN AFFINE ENERGY ON A HIERARCHY

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

$$\epsilon_h : \text{Part}(X) \rightarrow \mathbb{R}^+, \quad \text{Part}(X) \rightarrow \lambda C(P) + D(P), \quad (9)$$

where $D$ and $C$ are two functions on $\text{Part}(X)$, and $\lambda \in \mathbb{R}^+$.

Find the partition $P \in \text{Part}(X)$ which minimizes the affine energy $\epsilon_h$ 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 [2].

Definition 2.1: [2] For any $x \in \mathcal{H}$, $\mathcal{H}$ is a hierarchy of partitions of $X$, if $\mathcal{H}$ 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;
- $\forall s \in X : \{s\} \in \mathcal{H}$.

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

$$\mathcal{H}(x) = \{ y \in \mathcal{H} \mid y \subseteq x \},$$

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

Definition 2.2: [2], [3] A cut of $\mathcal{H}$ is a partition of $X$ whose elements are in $\mathcal{H}$.

We note $\text{Cut}(\mathcal{H})$ 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 2.3: [3] We say that $G : \text{Part}(X) \rightarrow \mathbb{R}^+$ is a multiscale energy, if there exists a function on the subsets of $X$ which we denote by $G$ such that:

$$G(P) = \sum_{P \in \text{Part}(X)} G(p), \quad \forall P \in \text{Part}(X); \quad (11)$$

We say that $G$ is subadditive if

$$G(S \cup R) \leq G(S) + G(R), \quad \forall S, R \in X, S \cap R = \emptyset; \quad (12)$$

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

From now on we assume that $\epsilon_h$ be an multiscale energy. For any $\lambda$, let $\Gamma_{\lambda}(\mathcal{H})$ be the cut of $\mathcal{H}$ minimizing $\epsilon_h$.

$$\Gamma_{\lambda}(\mathcal{H}) = \arg \min_{\Gamma \in \text{Cut}(\mathcal{H})} \epsilon_h(\Gamma). \quad (13)$$

Let $L_{\lambda}(\mathcal{H})$ the set of nodes of $\mathcal{H}$ which are locally optimal in $\mathcal{H}$ for the energy $\epsilon_h$,

$$L_{\lambda}(\mathcal{H}) = \{ x \in \mathcal{H} \mid \forall y \in \text{Cut}(\mathcal{H}), \epsilon_h(x) - \epsilon_h(y) \} \quad (14)$$

For any $x \in \mathcal{H}$, let $\Delta^*(x) = \{ \lambda \mid x \in L_{\lambda}(\mathcal{H}) \}$. The set $\Delta^*(x)$ represents the set of scales for which $x$ is locally optimal in $\mathcal{H}$ for the multiscale energy $\epsilon_h$. $\Delta^*(x)$ is an interval of the form $[a, +\infty)$.

Proposition 2.1: [2], [3] For any $x \in \mathcal{H}$,

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

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 2.4: [2] The persistent hierarchy obtained from $\mathcal{H}$ and $\epsilon_h$ is

$$\mathcal{H}^* = \{ X \in \mathcal{H} : \Delta^*(X) \neq \emptyset \}. \quad (15)$$

Remark 2.1: [2] On the persistent hierarchy $\mathcal{H}^*$ we have

$$\lambda^-(x) = \lambda^+(x^f), \quad (16)$$

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

Since $\epsilon_h$ is an multiscale energy, there exists two functions on the subsets of $X$ which we denote by $G_C$ and $G_D$ such that:

$$G(P) = \sum_{P \in \text{Part}(X)} G_C(p) + G_D(p), \quad (11)$$

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

$$\epsilon_x : \lambda \mapsto \lambda G_C(x) + G_D(x). \quad (17)$$

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

$$\epsilon_x^*(\lambda) = \epsilon_x(\Gamma_{\lambda}(\mathcal{H}(x))). \quad (18)$$

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

$$\epsilon_x^*(\lambda) = \epsilon_x(\lambda). \quad (19)$$
Proposition 2.2: [2], [3] The partial energies \( \varepsilon^*_x(\lambda) \) of the nodes of \( \mathcal{H} \) are related by the dynamic programming equation:

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

(20)

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

Proposition 2.3: [2], [3] For any \( x \in \mathcal{H} \),

\[
\varepsilon^*_x(\lambda) = \left\{ \sum_{s \in \mathcal{F}(x)} \varepsilon^*_s(\lambda), \lambda < \lambda^+(x); \varepsilon_x(\lambda), \lambda \geq \lambda^+(x) \right\},
\]

(21)

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

\[
\varepsilon_x(\lambda) = \sum_{s \in \mathcal{F}(x)} \varepsilon^*_s(\lambda).
\]

(22)

III. VIDEO SEGMENTATION ALGORITHM

If the measure \( \mathcal{H} \) is separable and strictly subadditive, then the 2D+time affine energy \( E_\lambda(C) = \lambda C(C) + D(C) \) is amultiscale energy, and \( C \) is strictly subadditive. 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 (8).

Let \( \bar{V} = \{ \bar{V}_m \}_{m \in [1, M]} \) be a 2D+time persistent hierarchy of video domain \( \Omega \times [T_1, T_2] \). In all what follows, we assume that the measure \( \mathcal{H} \) is separable and strictly subadditive, then \( E_\lambda(C) = \lambda C(C) + D(C) \) is amultiscale energy, and \( C \) is strictly subadditive, and the combining of the results of Propositions 2.1, 2.2 and 2.3 permit to construct an algorithm of video segmentation, that can compute a solution for the optimization problem (8) on \( \bar{V} \). The implementation of this video segmentation algorithm is based on four steps.

Video segmentation algorithm

**INPUT:** \( \bar{V} = \{ \bar{V}_m \}_{m \in [1, M]} \): 2D+time persistent hierarchy.

**Step 1:**
for \( m = 1 \) to \( M \)
do \{ compute \( E_m^* \) using the dynamic programming equation of Proposition 2.2 \} end for

**Step 2:**
for \( m = 1 \) to \( M \)
do \{ compute \( \lambda_m^* \) using the intersection between \( E_m \) and \( \sum_{s \in \mathcal{F}(m)} E_s^* \) (Proposition 2.3) \} end for

**Step 3:**
for \( m = 1 \) to \( M \)
do \{ compute \( \lambda_m^* \) using Proposition 2.1 \} end for

**Step 4:**
for \( m = 1 \) to \( M \)
if \( \lambda_m^* \leq \lambda \leq \lambda_m \)
do \{ Store \( \bar{V}_m \) \} end if

**OUTPUT:** \( \Gamma_\lambda(\bar{V}) = \{ \bar{V}_{\text{stored}} \} \) (Proposition 2.1).

IV. CONSTRUCTION OF 2D+TIME PERSISTENT HIERARCHY.

Now we propose an approach to constructing a 2D+time persistent hierarchy of \( \Omega \times [T_1, T_2] \). Our approach is based on the 2D-shapes of images [3], and Scale-invariant feature transform (SIFT) [4].

A. 2D hierarchy

1) Merging algorithm: Let \( S(f_n) \) be the tree of the shapes of image \( f_n = f(\cdot, t_n) \) of video sequence \( \{ f(\cdot, t_n) \}_{n \in \mathbb{N}} \), and \( P_n \) be a partition of \( f_n \) in \( S(f_n) \). We suppose here that the tree of the shapes have a finite number of elements, then \( P_n = \bigcup_{j=1}^J R_{n,j} \), where \( R_{n,j} \) are the regions of the partition \( P_n \). 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 \( S(f_n) \). We define a new partition \( P'_{n,j} \) by merging the regions \( \{B_1, ..., B_p, R_{n,j}\} \).

\[
P'_{n,j} = \{P_n \backslash \{B_1, ..., B_p, R_{n,j}\}\} \cup A
\]

(23)

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

Where \( \mathcal{E}_\lambda(P_n) = \lambda \mathcal{H}(P_n) + \int_{P_n} f(\bar{x}, t_n) - f(\bar{x}, t_n) \) d\( \bar{x} \)

merging algorithm

**INPUT:** \( P_0, A = \bigcup_{j=1}^J R_{0,j} \).

**Step 1:**
for \( j = 1 \) to \( J \)
do \{ compute \( \Delta \mathcal{E}(P_0, R_{0,j}) \) and insert \( R_{0,j} \) in a queue \( Q = \{q_1, j = 1, ..., J\} \) with priority \( \Delta \mathcal{E}(P_0, R_{0,j}) \) end for

**Step 2:**
Iterate the following procedure
if \( \Delta \mathcal{E}(P_n, q_i) > 0 \)
do \{ define \( P_{n+1,i} = \{P_n \backslash \{B_1, ..., B_p, q_i\}\} \cup A \) \} end if

We stop when no node \( q_i \) exists with \( \Delta \mathcal{E}(P_n, q_i) > 0 \).

**OUTPUT:** \( P_n = P_{n+f,n} \) is a partition.

The last partition obtained \( P_n \) determines by the father regions of the regions \( \{R_{n,j}\}_{j \in [1, J]} \).

2) Building of 2D hierarchy: For each image \( f_n \), we start with initial scale value \( \lambda_0 \), and a initial partition \( P_{0,n} = P_{0,n} \) of \( S(f_n) \). Let \( \bar{P}_{0,n} \) be the partition obtained by merging algorithm in \( P_{0,n} \). We continue iteratively the process of merging algorithm by minimizing the function \( \mathcal{E}_{\lambda_{k+1}}^{k+1} \lambda_{k+1} = 2\lambda_k \) on the partition \( \bar{P}_{k,n} \). The iterative process may be stopped either when the value of \( \lambda_k \) attains a maximum scale value \( \lambda_{k-1} \), finally we add the root partition \( \bar{P}_{k,n} = \Omega \).

Since, the partitions obtained \( \{P_{k,n}\}_{0 \leq k \leq K} \) are defened by shapes of the tree \( S(f_n) \), and any two shapes are either disjoint or nested, the partitions obtained \( \{\bar{P}_{k,n}\}_{0 \leq k \leq K} \) determine a 2D hierarchy of image \( f_n \).

B. 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 hierarchies defined by \( K \) partitions \( \{\bar{P}_{k,1}\}_{0 \leq k \leq K} \) and \( \{\bar{P}_{k,2}\}_{0 \leq k \leq K} \) obtained by merging algorithm, and we compute the keypoints \( \{p_{1,i}\}_{1 \leq i \leq K} \) and \( \{p_{2,i}\}_{1 \leq i \leq K} \) in the SIFT framework of image \( f_1 \) and \( f_2 \), and for all keypoint \( \{p_{1,i}\}_{1 \leq i \leq K} \), we search the matching keypoints in \( \{p_{2,i}\}_{1 \leq i \leq K} \). 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 \( \{\bar{P}_{k,1}\}_{0 \leq k \leq K} \) and \( \{\bar{P}_{k,2}\}_{0 \leq k \leq K} \):

- We compute the similar regions in \( \bar{P}_{0,1} \) and \( \bar{P}_{0,2} \) and we temporarily connect these similar regions.
Each non-similar region $R_{0,1,j} \in \mathcal{P}_{0,1}$ is characterized by a set $\{p_{1,i}\}_{1 \leq i \leq l_{0,j}}$ of keypoints $\{p_{1,i}\}_{1 \leq i \leq l_{0,j}}$. Let $\{p_{2,i}\}_{1 \leq i \leq l_{0,j}'}$ be the matching keypoints of keypoints $\{p_{1,i}\}_{1 \leq i \leq l_{0,j}}$ in image $f_2$. Let $\mathcal{M}(R_{0,1,j})$ be the set of matching regions of $R_{0,1,j}$ in $\mathcal{P}_{0,2}$. We note $\mathcal{S}(\mathcal{M}(R_{0,1,j}))$ the father regions of the set regions $\mathcal{M}(R_{0,1,j})$ in the 2D hierarchy $\{\mathcal{P}_{k,2}\}_{0 \leq k \leq K}$, and we define a new partition $\mathcal{P}_{2,2}'$ by replacing the set of matching regions $\mathcal{M}(R_{0,1,j})$ by $\mathcal{S}(\mathcal{M}(R_{0,1,j}))$ and we temporarily connect connect $R_{0,1,j}$ with $\mathcal{S}(\mathcal{M}(R_{0,1,j}))$.

We continue this process for computing the temporarily connected partitions $\mathcal{P}_{k,2}'$ for each partition level $k \in \{1, \ldots, K\}$. We apply 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 $\{\mathcal{P}_{k,n}\}_{1 \leq k \leq K}$ determine a 2D hierarchy of $\Omega$, and for each partition level $k \in \{1, \ldots, K\}$, the regions of $\{\mathcal{P}_{k,n}\}_{1 \leq n \leq N}$ are temporarily connected. So $\mathcal{V}_0 = \{\mathcal{P}_{k,n}\}_{1 \leq k \leq K; 1 \leq n \leq N}$ is a 2D+time hierarchy of $\Omega \times [T_1, T_f]$.

The nodes non-persistence of $\mathcal{V}_0$ bare not interested for the minimization of an multiscale energy $E_\lambda$ 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 a Greedy algorithm [3].

Fig. 1. Four consecutive frames of Dance and Monkey bar sequences.

Fig. 2. Comparing the accuracy and coherence of the proposed approach on the Dance sequence, to other approaches. First row: Hierarchical Graph-Based approach. Second row: Streaming Mean-shift Approach. Last row: Our approach.

V. EXPERIMENTAL RESULTS

In this section we present some experimental results obtained using the video segmentation procedure described. In fig 2 and 3, we compare our results of Monkey bar and Dance sequences (fig 1), against others on the leading methods that treat the video as a 3D space-time volume; Streaming Mean-Shift approach (SMS) [6], and Hierarchical Graph-Based (HGB) [7]. Fig 1 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 2 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.

VI. CONCLUSION

We have described a new method for segmentation method of video sequence using the minimization of an extension of Mumford-Shah Functional on a time+2D hierarchy. the idea is the writing our extension of Mumford-Shah functional as an affine energy of video partitions and computing a video segmentation by selecting a partition of video domain, witch minimize the affine energy.

REFERENCES