

# PARTITION LATTICE OPERATORS FOR EXTRACTION OF SEMANTIC VIDEO OBJECTS

DANIEL GATICA-PEREZ and MING-TING SUN  
*Department of Electrical Engineering, University of Washington.  
Box 352142. Seattle, WA, 981095, USA.*

and

CHUANG GU  
*Microsoft Corporation.  
One Microsoft Way. Redmond, WA 98052, USA.*

**Abstract.** We conceive the problem of multiple semantic video object (SVO) extraction as an issue of designing operators on a complete lattice of partitions. In this paper, we propose a framework based on accurate spatial partition generation and application of optimal extraction operators on the generated partitions. Under the framework, we introduce a spatio-temporal regional maximum likelihood operator for extraction purposes. Some theoretical properties of the operators are established. Experimental results show that our scheme is capable of successfully handling multiple SVOs in a variety of scenarios.

**Key words:** Semantic Video Object Extraction, Partition Lattice Operators.

## 1. Introduction

SVO extraction can be considered as a process of segmenting and tracking *arbitrary* collections of image regions (that correspond to objects in the real world) with *pixel-wise accuracy*. The task, crucial for the next generation of multimedia standards MPEG-4 and MPEG-7 [12], is formidable because SVOs are *human abstractions* that are not invariant, either in spatial features or in motion. Several approaches, based on different object representations (contours, regions, active meshes) have been recently proposed [2], [7], [11], [10].

When 2D regions are selected to represent an object, two factors define the quality of the extraction result: the precision of the spatial partition, and the selected tracking technique. On one hand, a good segmentation technique should preserve the contours of the scene objects, as human perception is sensitive to artifacts in borders. On the other hand, the tracking process is responsible for keeping an accurate SVO representation along time. Several methods consist of the computation of an initial object partition and its tracking by a *prediction-adjustment* process, in which the original partition is updated to generate the partitions at each time [3], [6], [10]. However, this process is not trivial if pixel-wise accuracy is required: noisy motion information, which constitutes the key factor of the procedure, often introduces inaccuracy and ambiguity in defining the boundaries of the video objects [7], [11]. Furthermore, some of these techniques unfortunately rely on heuristics, and/or cannot handle the case of

several objects.

This paper presents a different approach. We propose a general framework for 2D region-based SVO extraction based on spatial partition generation and the application of one or more extensive operators in the lattice of partitions. We are interested in developing a systematic approach in which such operators can be defined, their properties can be analyzed, and that allows for the extraction of multiple SVOs from natural video sequences.

The rest of the paper is organized as follows. Section 2 presents an overview of our methodology. Section 3 introduces the partition lattice operators for SVO extraction. Section 4 shows results for several MPEG-4 test sequences. Section 5 provides some concluding remarks.

## 2. Proposed Approach

A complete lattice of partitions is an appropriate morphological framework to analyze segmentation problems [15], [14]. We start by reviewing such a notion. Given a space  $\mathcal{E}$  and its power set  $\mathcal{P}(\mathcal{E})$ , a *partition* of  $\mathcal{E}$  is a mapping  $P : \mathcal{E} \rightarrow \mathcal{P}(\mathcal{E})$  such that  $\forall x, y \in \mathcal{E}$ , (i)  $x \in P(x)$ , and (2)  $P(x) = P(y)$  or  $P(x) \cap P(y) = \emptyset$ .  $P(x)$  is called the *zone* or *region* of  $P$  that contains  $x$ . It can be proved that the set of all partitions of  $\mathcal{E}$  constitutes a complete lattice, denoted by  $\Pi$ , where the partial ordering relationship is defined as  $P_i \leq P_j \iff P_i(x) \subseteq P_j(x), \forall x \in \mathcal{E}, P_i, P_j \in \Pi$ . In this case,  $P_i$  is said to be *finer* than  $P_j$ . The infimum of a set of partitions  $\{P_i, i \in \mathcal{I}\}$  is defined as  $(\bigwedge_i P_i)(x) = \bigcap_i P_i(x) \forall x \in \mathcal{E}$ , i.e., it corresponds to the partition made of the intersections of all the regions in the original set of partitions. Additionally, the supremum of a set  $\{P_i\}$  is given by  $(\bigvee_i P_i)(x) = \bigcap \{B : B = \bigcup_i \bigcup_{y \in B} P_i(y), x \in B, B \in \mathcal{P}(\mathcal{E})\}$  which is the finest partition that is larger than each of the individual  $P_i$ . For the two-partition case,  $(P_i \vee P_j)(x) = (P_i \vee P_j)(y)$  if  $P_i(x) = P_i(y)$  or  $P_j(x) = P_j(y)$ . Finally, the least and greatest elements of  $\Pi$  correspond to the finest partition  $P_O$  and the coarsest partition  $P_I$ , such that  $P_O(x) = x$  and  $P_I(x) = \mathcal{E}$  for all  $x \in \mathcal{E}$ .

For purposes of indexing of the zones of a partition, it is convenient to use the following notation:  $P = \{R_i, i \in \mathcal{I}\}$ , where  $R_i = \bigcup x \in \mathcal{E}$  such that  $P(x) = R_i$ .

The extraction of the SVOs of a scene corresponds to one special case of partition of the image support. This can be defined as follows.<sup>1</sup>

**Definition 1** Let  $\mathbf{I} = \{\mathbf{I}^t \mid t \in \mathbf{Z}\}$  be a multivalued image sequence, with domain  $\mathcal{E} = \mathcal{D}(\mathbf{I}^t) \subset \mathbf{Z}^2$ . Let  $P^t = \{R_i^t, i \in \{1, \dots, N\}\}$  denote a partition of  $\mathcal{E}$  at time  $t$ . The  $j$ -th Semantic Video Object of the scene depicted in  $\mathbf{I}$  (consisting of  $M$  objects) is defined by  $SVO_j = \{SVO_j^t\}$ , where

$$SVO_j^t = \bigcup_{i=1}^{N_j^t} R_i^t \quad (1)$$

In MPEG-4 terminology,  $SVO_j^t$  represents the  $j$ -th Video Object Plane (VOP) at time  $t$ , each composed of  $N_j^t$  regions of  $P^t$  ( $\sum_j N_j^t = N$ ). Eq.

---

<sup>1</sup> Multivalued images and random variables are both denoted by bold letters. The meaning should be clear from the context.

1 naturally allows for the definition of multiple SVOs. The associated partition of SVOs at time  $t$ , denoted by  $P_{SVO}^t$ , is the collection

$$P_{SVO}^t = \{SVO_j^t, j \in \{1, \dots, M\}\} \quad (2)$$

We propose to achieve SVO extraction by (1) generating, at each time instant, spatial partitions  $P^t$  that do not depend on inaccurate motion information and that preserve the true object contours, so that the frontiers between objects can be discerned even though they are of similar color, and (2) finding optimal homomorphisms between the generated partitions and the set of SVOs in the scene. Therefore, we claim that SVO tracking can be formalized as a process of applying operators  $\{\psi^t()\}$  in the lattice of partitions [14], [15]. These operators can be designed by specifying spatio-temporal statistical criteria.

In addition, we consider the interactive introduction of semantics as essential, such that one or more of the following functions can be implemented by a user: initial definition of the SVOs, creation of a multiview representation, correction of the automatic results, or specification of context. Some typical works in this area include [2] and [7].

SVO extraction is then defined as a *combined process of spatial partition generation and subsequent application of partition lattice operators, with user intervention at (possibly) different instants*.

For the partition generation stage, we have previously developed a four-band (color+intensity edges) morphological multivalued spatial segmentation method that improves the contour localization properties of the traditional watershed techniques [5]. The following section concentrates on the formulation of the partition operators.

### 3. Partition Lattice Operators for Semantic Video Object Extraction

Some basic partition operators for segmentation were originally proposed in [14]. More recently, a work that pointed out the connection between region merging algorithms and connected operators was presented in [4].

In our approach, once a partition  $P^t$  is generated at each time  $t$ , the problem becomes the *construction* of  $P_{SVO}^t$  from  $P^t$  by introducing temporal information that allows the implementation of the tracking function. This information is represented by a *temporal reference SVO partition set*, denoted by  $TR$ , composed of the SVO partitions of the scene at different time instants, i.e., partitions that correspond to different scene views. The partition reference set is then expressed as  $TR = \{P_{SVO}^{t_k}, k \in \{1, \dots, K\}\}$ . For example, if  $K = 1$ , and  $t_K = t - 1$ , then  $TR = \{P_{SVO}^{t-1}\}$ , which means that the generation of the current SVO partition will depend on information provided by the previous one. If  $K > 1$ , the decision for the construction of the current SVO partition will include information from multiple instances. Note that the partitions in  $TR$  can be computed either by off-line user interaction or automatically as part of the extraction process. We now define a SVO extraction operator.

**Definition 2** *Let  $\Pi$  be the complete lattice of partitions of  $\mathcal{E} = \mathcal{D}(\mathbf{I}^t)$ . Let  $P^t \in \Pi$ , and  $TR = \{P_{SVO}^{t_1}, \dots, P_{SVO}^{t_K}\}$ . An SVO extractor operator is a mapping*

$\psi_{TR}^t : \Pi \rightarrow \Pi$ , so that

$$P_{SVO}^t = \psi_{TR}^t(P^t) \quad (3)$$

The explicit dependence on  $TR$  will be usually omitted as this set is used as a reference. With this formulation, several tracking schemes can be formulated: monoview ( $k = 1$ ) or multiview ( $k > 1$ ); causal ( $t_k < t$ ) or non-causal ( $t_k > t$ ). The superscript in the operators notation means that they can be time-dependent.

As we mentioned, we have assigned the accurate extraction of region boundaries to the partition generation phase. As a result, the extraction operators  $\{\psi^t\}$  can be thought of as a classification mechanism, that assigns each region  $R_i^t \in P^t$  to the appropriate SVO, so that no new spatial contours are introduced in the partition  $P_{SVO}^t$ . This concept is illustrated in Fig.1. In fact, all possible operators that can be designed with this idea in mind satisfy the following property.

**Property 1** *The operators  $\{\psi^t\}$  are extensive.*

*Proof.* It directly follows from Eqs. 1 and 2.

This property implies a relation between this class of operators and connected operators [4], [9]. In addition, we would like the operators  $\{\psi^t\}$  not to be injective. From the practical point of view, this would represent some robustness in the extraction operation, by allowing several partitions  $\{P^t\}$ , all comparable by the ordering relation, to be mapped to the same object partition  $P_{SVO}^t$ . Finally, idempotence represents another desired property, as it would imply that the extraction would be done in one single step.



Fig. 1. SVO extraction operator. *Hand* sequence. (a) Original image  $\mathbf{I}^1$ . (b) Partition  $P^1$  (segmentation model superimposed); (c) SVO partition  $P_{SVO}^1 = \psi(P^1)$  (original image superimposed).

In the following section, we present one operator for the case in which the reference set is given by  $TR = \{P_{SVO}^{t-1}\}$ , and  $\psi$  is non-time-adaptive.

### 3.1. PARTITION OPERATOR BASED ON REGIONAL MAXIMUM LIKELIHOOD.

The design of the partition operators can be formulated in terms of an optimality criterion to be satisfied. Note that, from the statistical point of view, SVO extraction (as has been formulated here) represents a process of assigning the regions of  $P^t$  to a given class, namely the objects in the scene; from the algebraic point of view, it corresponds to the design of extensive partition operators. We initially propose a partition operator  $\psi_j(P^t)$  to construct *each*  $SVO_j^t$  from the partition  $P^t$  using the previous SVO partition ( $P_{SVO}^{t-1}$ ) as reference,

$$P_{SVO_j}^t = \psi_j(P^t)$$

where  $P_{SVO_j}^t$  is the partition that divides the image support into the j-th SVO and the rest of the scene. The generation of  $P_{SVO}^t$  is then straightforward.

Let  $V_{\mathbf{I}^t, \mathbf{I}^{t-1}} : \mathcal{P}(\mathcal{E}) \rightarrow \mathbf{Z}^2$  be the mapping that computes a region motion vector, assuming a pure translational model, using the image sequence  $\mathbf{I}^t$  at times  $t$  and  $t-1$ , so that  $V_i^t = V_{\mathbf{I}^t, \mathbf{I}^{t-1}}(R_i^t)$  denotes the motion vector computed for the region  $R_i^t \in P^t$ . Additionally, let  $[X]_h$  represent the translated version of  $X \subset \mathbf{Z}^2$  by  $h \in \mathbf{Z}^2$ :  $[X]_h = \{x + h \mid x \in X\}$ . The region attribute that will be used to construct the SVO partition at each time  $t$ , using the temporal reference  $P_{SVO}^{t-1} = \{SVO_j^{t-1}, j \in \{1, \dots, M\}\}$  is defined as follows.

**Definition 3** Given a partition  $P^t$ , and the SVO partition  $P_{SVO}^{t-1}$ , the normalized overlapped area between the  $i$ -th region  $R_i^t \in P^t$  and the  $j$ -th SVO is given by:

$$noa_{ij}^t = \frac{\text{card}([R_i^t]_{V_i^t} \cap SVO_j^{t-1})}{\text{card}(R_i^t)} \quad (4)$$

This measure takes values between zero (no overlapping) and one ( $R_i^t \subseteq SVO_j^{t-1}$ ), and will decide for the assignment of each region in  $P^t$  to the corresponding SVO. Obviously, each  $R_i^t \in P^t$  belongs either to the  $j$ -th SVO or to any other SVO in the scene depicted in the image sequence. In hypothesis testing terms,

$$H_0 : R_i^t \subseteq SVO_j^t \quad ; \quad H_1 : H_0^c \quad (5)$$

The normalized overlapped area can be modeled as a continuous random variable **noa**, taking values  $noa$  in  $[0,1]$  (we drop the index  $t$  in what follows to simplify the notation). Let  $svo_j, j = 1, \dots, M$  represent the  $j$ -th possible class (i.e. the  $j$ -th SVO), with prior probabilities  $\Pr(svo_j)$ , and let  $svo_j^c$  denote the set of all classes except the  $j$ -th one, which implies  $\Pr(svo_j^c) = 1 - \Pr(svo_j)$ . With this setting,  $\Pr(svo_j|noa)$  and  $\Pr(svo_j^c|noa)$  represent the a posteriori conditional probabilities that correspond to  $H_0$  and  $H_1$ , respectively. We use the Maximum a Posteriori (MAP) criterion to map each region to an SVO [13]:

$$\Pr(svo_j|noa) \underset{H_0}{\overset{H_1}{\gtrless}} \Pr(svo_j^c|noa) \quad (6)$$

such that the hypothesis  $H_x$  that is chosen is the one that has a larger a posteriori probability. Applying Bayes theorem on both sides of the expression and rearranging terms,

$$\frac{p(noa|svo_j)}{p(noa|svo_j^c)} \underset{H_0}{\overset{H_1}{\gtrless}} \frac{\Pr(svo_j^c)}{\Pr(svo_j)} \quad (7)$$

where  $p(noa|svo_j)$  represents the class-conditional probability density function. For the two-object case, we can assume equal priors ( $\Pr(svo_j) = \Pr(svo_j^c)$ ), as

foreground and background video objects may have any size and shape, and the expression reduces to the maximum likelihood criterion

$$L(noa) \equiv \frac{p(noa|svoj)}{p(noa|svoj^c)} \underset{H_0}{\overset{H_1}{>}} 1 \quad (8)$$

For the cases of larger number of objects, however, the exact expression is Eq. 7. Let  $k^t$  denote the ratio  $\Pr(sv oj^c)/\Pr(sv oj)$ . We propose to model the class-conditional probability density functions by exponential distributions:

$$p(noa|svoj^c) = \lambda_1 e^{-\lambda_1 noa} u(noa) \quad ; \quad p(noa|svoj) = \lambda_2 e^{-\lambda_2(1-noa)} u(1-noa)$$

where  $u(x)$  designates the step function. These distributions approximately model the real data: due to segmentation errors,  $p(noa|svoj)$  should be highly concentrated around  $noa = 1$ , and rapidly decay as  $noa \rightarrow 0$ . The dual situation holds for  $p(noa|svoj^c)$ . In addition, the parameter values  $\lambda_i$  should make the conditional probabilities outside the interval  $[0, 1]$  negligible. The problem has been reduced to finding an optimal threshold for  $noa$ ,

$$noa \underset{H_0}{\overset{H_1}{>}} \frac{\lambda_2 - \ln(\lambda_2/k^t \lambda_1)}{\lambda_1 + \lambda_2} = T_{noa} \quad (9)$$

We can now write an expression for the proposed partition operator:

$$P_{SVO_j}^t = \psi_j(P^t) = \{SVO_j^t, \mathcal{E} \setminus SVO_j^t\} \quad (10)$$

where  $A \setminus B$  denotes set difference and

$$SVO_j^t = \bigcup_i R_i^t \text{ such that } noa_{ij}^t \geq T_{noa} \quad (11)$$

The parameters  $\lambda_i$  and  $k^t$  can be estimated from the actual data. However, if we assume symmetry between the exponential distributions ( $\lambda_1 = \lambda_2$ ), and  $\lambda_i \gg k^t$ , the expression for the optimal threshold can be further simplified and approximated as:

$$T_{noa} = \frac{\lambda_2 - \ln(\lambda_2/\lambda_1)}{\lambda_1 + \lambda_2} + \frac{\ln k^t}{\lambda_1 + \lambda_2} \approx \frac{1}{2} \quad (12)$$

This analysis shows that  $\psi_j$ , under the described assumptions, is equivalent to a tracking algorithm recently reported in [7] for the two-SVO case.

To extract the  $M$  SVOs present in the scene,  $\psi_j$  should be applied  $M - 1$  times (the  $M - th$  SVO is always selected as the scene background). Finally,  $P_{SVO}^t$  can be directly generated from the set of partitions  $\{\psi_j(P^t)\}_{j=1}^{M-1}$ , by defining a partition operator  $\psi_{RML}()$  for *regional maximum likelihood*:

$$P_{SVO}^t = \psi_{RML}(P^t) = \bigwedge_{j=1}^{M-1} \psi_j(P^t) = \bigwedge_{j=1}^{M-1} P_{SVO_j}^t \quad (13)$$

Some properties of this operator can be established (the same applies to  $\psi_j$ , as it is equivalent to  $\psi_{RML}$  for  $M = 2$  in Eq. 13).

**Property 2**  $\psi_{RML}$  has the following properties: (i) neither increasing nor decreasing, (ii) idempotent, (iii) not injective, (iv) not invertible, (v) not a morphological filter.

*Proof (i).* (Counterexample). Let  $P_i^t$  be a partition of  $\mathcal{E}$  that consists of three regions, labeled  $B$  (background),  $H$  (head), and  $S$  (shoulders), respectively. Let  $P_{i'}^t$  be another partition that consists of two regions,  $H$  and  $E = B \cup S$  (erroneously merged regions). By construction,  $P_i^t \leq P_{i'}^t$ . Additionally, assume that there is no motion and that  $P_{SVO}^{t-1}$  is correctly composed of the background  $B$  and the object  $O = H \cup S$ . Applying  $\psi_{RML}$  to the two partitions,  $\psi_{RML}(P_i^t) = P_{SVO}^{t-1}$ , and  $\psi_{RML}(P_{i'}^t) = P_{i'}^t$ , but obviously  $P_{SVO}^{t-1} \not\leq P_{i'}^t$ , so  $\psi_{RML}$  is not increasing. Similarly, it can be proved that  $\psi_{RML}$  is not decreasing.

(ii)  $\psi_{RML}(\psi_{RML}(P^t)) = \psi_{RML}(P_{SVO}^t)$ . But  $P_{SVO}^t$  is already the partition of SVOs. A further classification process simply assigns every  $SVO_j$  to itself, i.e.  $\psi_{RML}(P_{SVO}^t) = P_{SVO}^t$ .

(iii) Using the counterexample in (i),  $\psi_{RML}(P_i^t) = \psi_{RML}(P_{SVO}^{t-1}) = P_{SVO}^{t-1}$ , so  $P_i^t$  and  $P_{SVO}^{t-1}$  map to the same partition under  $\psi_{RML}$ , which shows that the operator is not injective. In general, for a set of partitions  $\{P_1^t \leq \dots \leq P_n^t\}$ , the equality  $\psi_{RML}(P_1^t) = \dots = \psi_{RML}(P_n^t)$  will hold.

(iv) Follows from (iii).

(v) Remember that a lattice operator is called a morphological filter iff it is idempotent and increasing. The result immediately follows from (i).

### 3.2. STATISTICAL VALIDATION OF THE PARTITION OPERATOR

To justify the assumptions in the previous subsection, we performed statistical tests on several MPEG-4 video sequences. Indeed, we found that the exponential, symmetrical distribution assumption adequately represents the data. In Table I, we show the ML estimates for the parameters  $\lambda_i$ , for the two-SVO case (foreground object and background). Additionally, the priors  $\Pr(svo_j)$  at each time  $t$  are estimated from the relative sizes of the SVOs at the previous frame of the video sequence, so that:

$$\hat{k}^t = \Pr(svo_j^c) / \Pr(svo_j) = \text{card}(\mathcal{E} \setminus SVO_j^{t-1}) / \text{card}(SVO_j^{t-1})$$

TABLE I  
Estimated parameters for MPEG-4 sequences

Sequence	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{k}^0$	$\hat{T}_{noa}$
Bream	142.21	113.89	2.44	0.45
Foreman	75.17	79.51	1.94	0.51
Hand	97.64	78.83	4.25	0.44

Table I also shows the initial values of  $\hat{k}^t$ . It is observed that the assumption that  $\lambda_i \gg k^t$  also holds, even for small objects, and that the estimated optimum threshold  $\hat{T}_{noa}$  is actually close to the approximated value. This fact validates the direct use of  $1/2$  as the value of  $T_{noa}$ , which reduces the computational complexity. It is also pointed out that  $\psi_{RML}$  can tolerate SVO size changes.

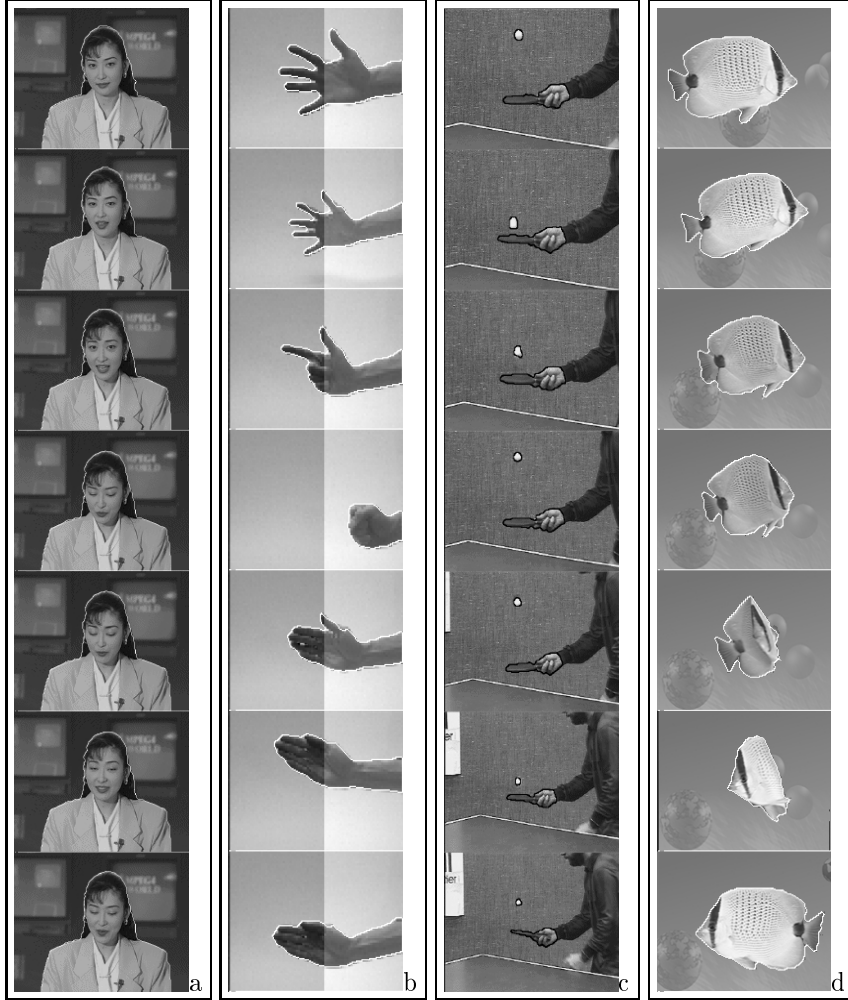


Fig. 2. SVO extraction. (a) *Akiyo*. (b) *Hand*. (c) *Tennis*. (d) *Bream*.

#### 4. Results

Our framework is integrated in an SVO extraction system whose minimum-user interaction model was presented in [5], and consists of four steps:

1. SVO structure definition. A user-defined  $P_{SVO}^0$  is generated from  $\mathbf{I}^0$ .
2. SVO computation by generation of a partition  $P^t \forall t$ .
3. SVO tracking by application of our partition operator,  $\hat{P}_{SVO}^t = \psi^t(P^t, P_{SVO}^{t-1})$ .
4. SVO postprocessing to refine the object partitions,  $\hat{P}_{SVO}^t \rightarrow P_{SVO}^t$ .

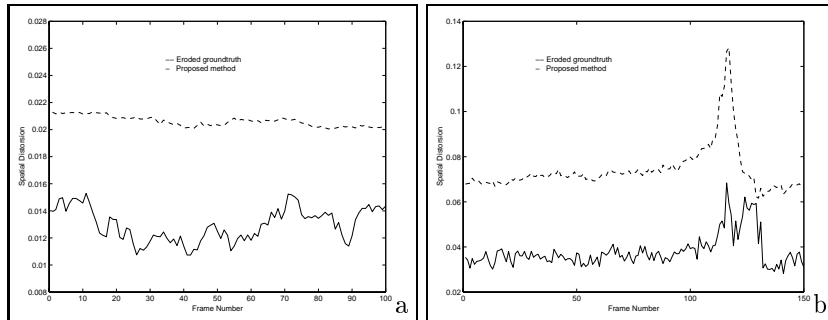


Fig. 3. Generated SVO Spatial Distortion. (a) *Akiyo*, (b) *Bream*. The distortion introduced by the erosion of the ground truth by a  $3 \times 3$  structuring element is also shown.

Extraction results for various scenarios are shown in Fig. 2<sup>2</sup>. In all cases, we superimpose  $P_{SVO}^t$  on  $I^t$ . Fig. 2(a) shows the result for the *Akiyo* sequence for two user-defined SVOs: *Akiyo* and *Background*. Even though they have some adjacent regions of similar color, our methodology generated precise SVO contours. Fig. 2(b) illustrates a two-SVO gesture image sequence [1]. The hand presents fast, articulated motion and shades, and the scene has a significant change of illumination. As a third example, the result obtained with the *Tennis* sequence, divided into three SVOs, is shown in Fig. 2(c). The sequence has been correctly partitioned. Finally, the result obtained with the *Bream* sequence, that presents object deformable motion and global camera motion, is shown in Fig. 2(d). In summary, our methodology performs well for different types of object and camera motion.

The computational complexity of our method is low, and adequate for semi-automatic SVO extraction. When fast motion estimation is used, the extraction takes around three seconds/frame in QCIF color images, on an SGI Octane computer; this figure could be significantly reduced by code optimization. Full motion estimation provides the best SVO extraction results at the expense of increasing the processing time, and might be required when tracking tiny objects with large motion.

Objective evaluation of our methodology can be performed for those sequences for which a ground truth is available. The MPEG-4 group has proposed figures for spatial distortion evaluation [16]. In Fig. 3 we present the results obtained for the *Akiyo* and *Bream* test sequences. To provide an idea of the degree of accuracy of the generated SVO partitions, the spatial distortion computed between the ground truth and a  $3 \times 3$ -eroded version of itself, that approximately peels off the ideal SVO partition by one pixel, is also presented, and confirms the obtained quality.

The main limitations of the proposed method arise when extracting SVOs in (i) highly cluttered scenes where the colors of different SVOs are similar, which introduces segmentation errors, and (ii) sequences in which newly uncovered regions have no matches in the previous frame, which produces tracking errors.

<sup>2</sup> Test video sequences are available at <http://hitl.washington.edu/people/danielgp>

We are currently extending our methodology to a multiview SVO representation (i.e., when the partition reference set  $TR$  is composed of more than one SVO partition) to address these problems.

## 5. Conclusions

We described a methodology for multiple SVO extraction based on object contour-preserving spatial partition generation and application of extensive spatio-temporal partitions operators. The use of the partition lattice framework for SVO extraction allows for the modeling of various tracking schemes and leads to the development of optimal algorithms. We have illustrated this with a regional maximum likelihood operator. Experimental results for a variety of real situations in natural video sequences have verified its effectiveness.

## Acknowledgements

D.G.P. thanks the support from the National University of Mexico and the Fulbright-CONACyT scholarship program.

## References

1. M. J. Black, and A. Jepson. Eigenttracking: Robust matching and tracking of articulated objects using a view-based representation. *Int. J. of Comp. Vis.*, 26(1), pp.63-84, 1998
2. P. Correia, and F. Pereira. The role of analysis in content-based video coding and indexing. *Signal Processing*, 66(2):125-142, April 1998.
3. V. Garcia-Garduno. Une approche de compression orientee-objets par suivi de segmentation basee mouvement. Ph.D. dissertation, Universite de Rennes I, France, 1995.
4. L. Garrido, P. Salembier and D. Garcia. Extensive operators in partition lattices for image sequence analysis. *Signal Processing*, 66(2):157-180, April 1998.
5. D. Gatica-Perez, M. T. Sun and C. Gu. Semantic video object extraction based on backward tracking of multivalued watershed. In *Proc. of the IEEE ICIP*, Kobe, Oct. 1999.
6. C. Gu. Multivalued Morphology and segmentation based coding. Ph.D. dissertation, LTS-EPFL, <http://ltswww.epfl.ch/staff/gu.html>, 1995.
7. C. Gu and M.-C. Lee. Semantic Video Object Tracking Using Region-based Classification. In *Proc. of the IEEE ICIP*, pp. 643-647, Chicago, Oct. 1998.
8. H.J.A.M. Heijmans. *Morphological Image Operators*. Academic Press, 1994.
9. H.J.A.M. Heijmans. Connected Morphological Operator for Binary Images. Report PNA-R9708, CWI, Amsterdam, April, 1997.
10. *IEEE Transactions on Circ. and Syst. for Video Tech.*, Special Issue on Segmentation, Description, and Retrieval of Video Content. Vol. 8, No. 5, September 1998.
11. F. Marques and J. Llach. Tracking of generic objects for video object generation. In *Proc. of the IEEE ICIP*, pp. 628-632, Chicago, Oct. 1998.
12. Microsoft Windows Media Player<sup>TM</sup>, in <http://www.microsoft.com/windows/windowsmedia/en/default.asp>, 1999.
13. A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 1993.
14. J. Serra. *Image Analysis and Mathematical Morphology, Vol. II: Theoretical Advances*. Academic Press, 1988.
15. J.C. Simon. *Patterns and Operators: the Foundations of Data Representation*. North Oxford Academic, 1986.
16. M. Wollborn and R. Mech. Refined procedure for objective evaluation of video object generation algorithms. Doc. ISO/IEC JTC1/SC29/WG11 M3448, March 1998.