Chapter 12
MPEG Video Coding II
— MPEG-4, 7 and Beyond

12.1 Overview of MPEG-4

12.2 Object-based Visual Coding in MPEG-4

12.3 Synthetic Object Coding in MPEG-4

12.4 MPEG-4 Object types, Profile and Levels

12.5 MPEG-4 Part10/H.264

12.6 MPEG-7

12.7 MPEG-21

12.8 Further Exploration

12.1 Overview of MPEG-4

- **MPEG-4**: a newer standard. Besides compression, pays great attention to issues about user interactivities.
- MPEG-4 departs from its predecessors in adopting a new **object-based coding**:
  - Offering higher compression ratio, also beneficial for digital video composition, manipulation, indexing, and retrieval.
  - Figure 12.1 illustrates how MPEG-4 videos can be composed and manipulated by simple operations on the visual objects.
- The bit-rate for MPEG-4 video now covers a large range between 5 kbps to 10 Mbps.

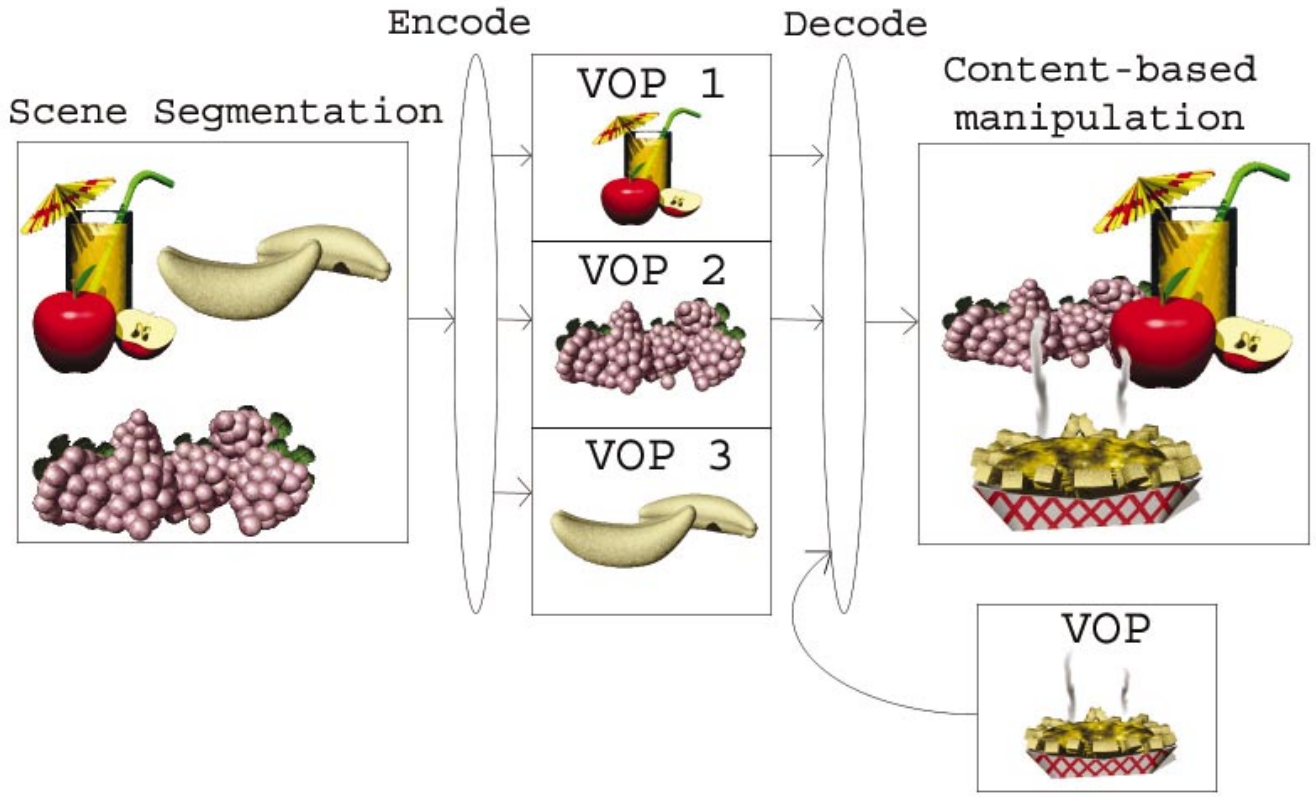

Fig. 12.1: Composition and Manipulation of MPEG-4 Videos.

Overview of MPEG-4 (Cont'd)

- MPEG-4 (Fig. 12.2(b)) is an entirely new standard for:

  (a) Composing media objects to create desirable audiovisual scenes.

  (b) Multiplexing and synchronizing the bitstreams for these media data entities so that they can be transmitted with guaranteed Quality of Service (QoS).

  (c) Interacting with the audiovisual scene at the receiving end — provides a toolbox of advanced coding modules and algorithms for audio and video compressions.

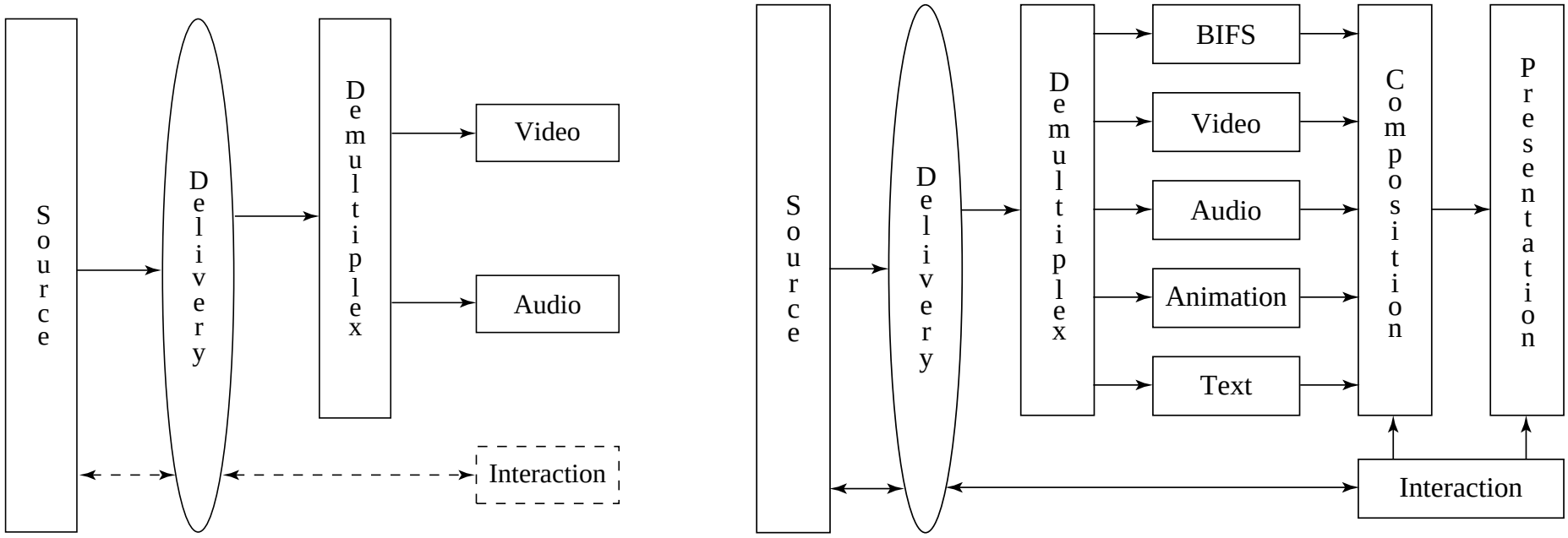

Fig. 12.2: Comparison of interactivities in MPEG standards: (a) reference models in MPEG-1 and 2 (interaction in dashed lines supported only by MPEG-2); (b) MPEG-4 reference model.

Overview of MPEG-4 (Cont'd)

- The hierarchical structure of MPEG-4 visual bitstreams is very different from that of MPEG-1 and -2, it is very much video object-oriented.

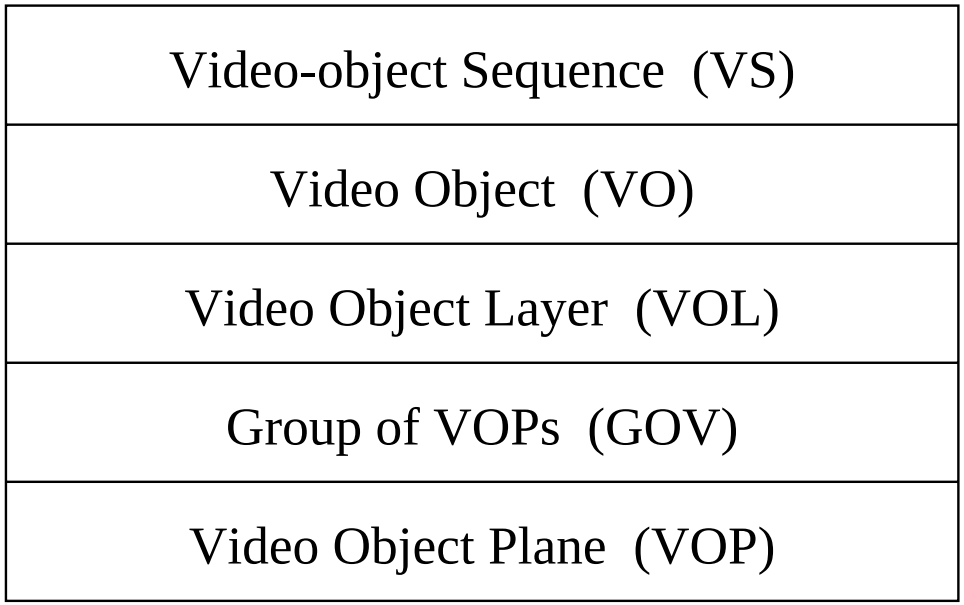

Fig. 12.3: Video Object Oriented Hierarchical Description of a Scene in MPEG-4 Visual Bitstreams.

Overview of MPEG-4 (Cont'd)

1. **Video-object Sequence** (VS) — delivers the complete MPEG-4 visual scene, which may contain 2-D or 3-D natural or synthetic objects.
2. **Video Object** (VO) — a particular object in the scene, which can be of arbitrary (non-rectangular) shape corresponding to an object or background of the scene.
3. **Video Object Layer** (VOL) — facilitates a way to support (multi-layered) scalable coding. A VO can have multiple VOLs under scalable coding, or have a single VOL under non-scalable coding.
4. **Group of Video Object Planes** (GOV) — groups Video Object Planes together (optional level).
5. **Video Object Plane** (VOP) — a snapshot of a VO at a particular moment.

12.2 Object-based Visual Coding in MPEG-4

VOP-based vs. Frame-based Coding

- MPEG-1 and -2 do not support the VOP concept, and hence their coding method is referred to as **frame-based** (also known as **Block-based coding**).

- Fig. 12.4 (c) illustrates a possible example in which both potential matches yield small prediction errors for block-based coding.

- Fig. 12.4 (d) shows that each VOP is of arbitrary shape and ideally will obtain a unique motion vector consistent with the actual object motion.

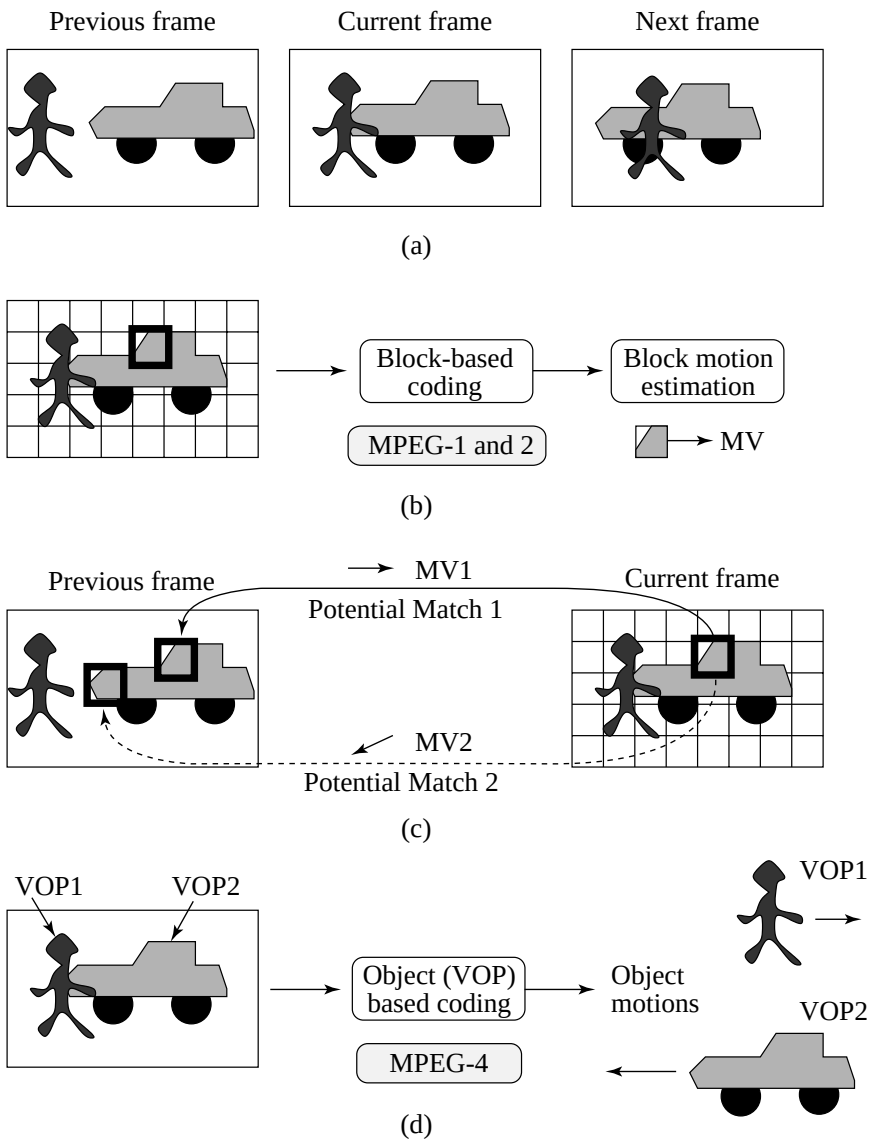

Fig. 12.4: Comparison between Block-based Coding and Object-based Coding.

VOP-based Coding

- MPEG-4 VOP-based coding also employs the Motion Compensation technique:
  - An Intra-frame coded VOP is called an **I-VOP**.
  - The Inter-frame coded VOPs are called *P-VOP*s if only forward prediction is employed, or *B-VOP*s if bi-directional predictions are employed.
- The new difficulty for VOPs: may have arbitrary shapes, shape information must be coded in addition to the texture of the VOP.

  Note: *texture* here actually refers to the visual content, that is the gray-level (or chroma) values of the pixels in the VOP.

VOP-based Motion Compensation (MC)

- MC-based VOP coding in MPEG-4 again involves three steps:

  (a) Motion Estimation.

  (b) MC-based Prediction.

  (c) Coding of the prediction error.

- Only pixels within the VOP of the current (Target) VOP are considered for matching in MC.

- To facilitate MC, each VOP is divided into many macroblocks (MBs). MBs are by default 16×16 in luminance images and 8×8 in chrominance images.

- MPEG-4 defines a rectangular *bounding box* for each VOP (see Fig. 12.5 for details).

- The macroblocks that are entirely within the VOP are referred to as **Interior Macroblocks**.

  The macroblocks that straddle the boundary of the VOP are called **Boundary Macroblocks**.

- To help matching every pixel in the target VOP and meet the mandatory requirement of rectangular blocks in transform codine (e.g., DCT), a pre-processing step of *padding* is applied to the Reference VOPs prior to motion estimation.

  Note: Padding only takes place in the Reference VOPs.

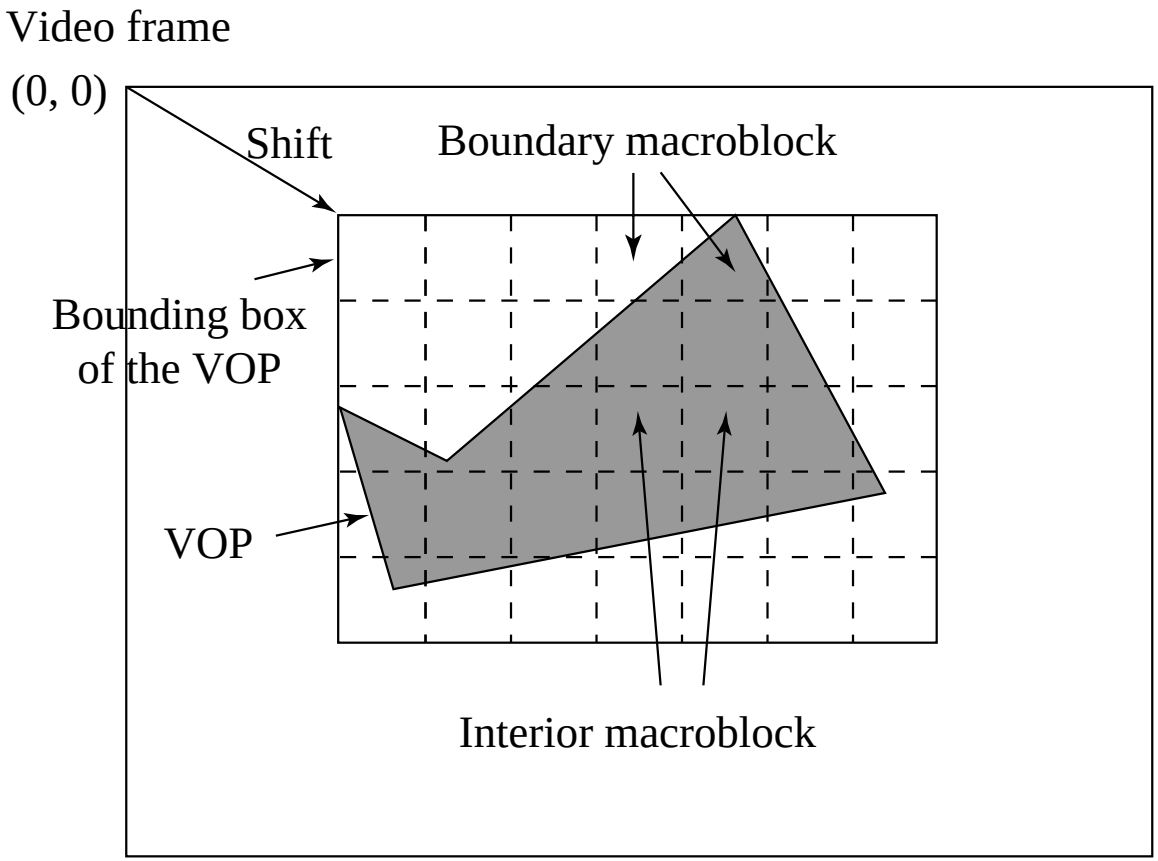

Fig. 12.5: Bounding Box and Boundary Macroblocks of VOP.

I. Padding

- For all Boundary MBs in the Reference VOP, *Horizontal Repetitive Padding* is invoked first, followed by *Vertical Repetitive Padding*.

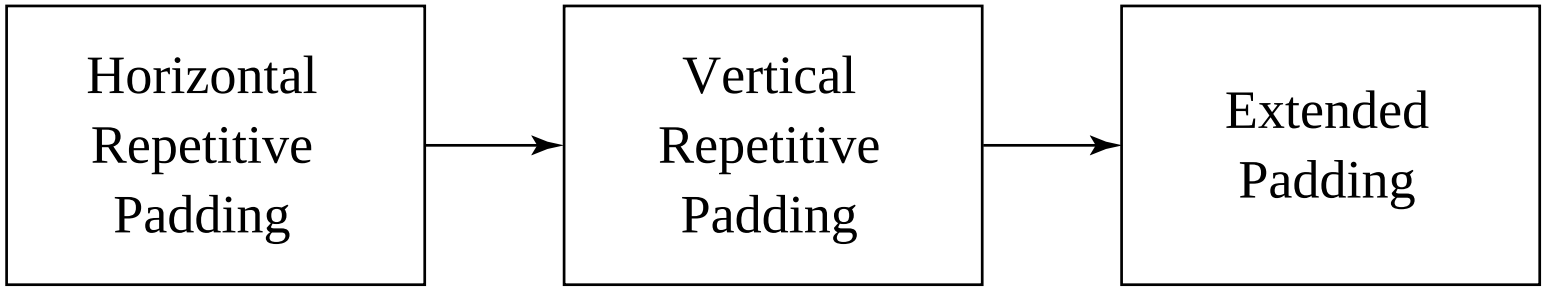

Fig. 12.6: A Sequence of Paddings for Reference VOPs in MPEG-4.

- Afterwards, for all **Exterior Macroblocks** that are outside of the VOP but adjacent to one or more Boundary MBs, *extended padding* will be applied.

Algorithm 12.1 Horizontal Repetitive Padding:

```
begin
   for all rows in Boundary MBs in the Reference VOP
     if ∃ (boundary pixel) in the row
       for all interval outside of VOP
         if interval is bounded by only one boundary pixel b
           assign the value of b to all pixels in interval
         else    // interval is bounded by two boundary pixels b1 and b2
           assign the value of (b1 + b2)/2 to all pixels in interval
end
```

- The subsequent Vertical Repetitive Padding algorithm works in a similar manner.

Example 12.1: Repetitive Paddings

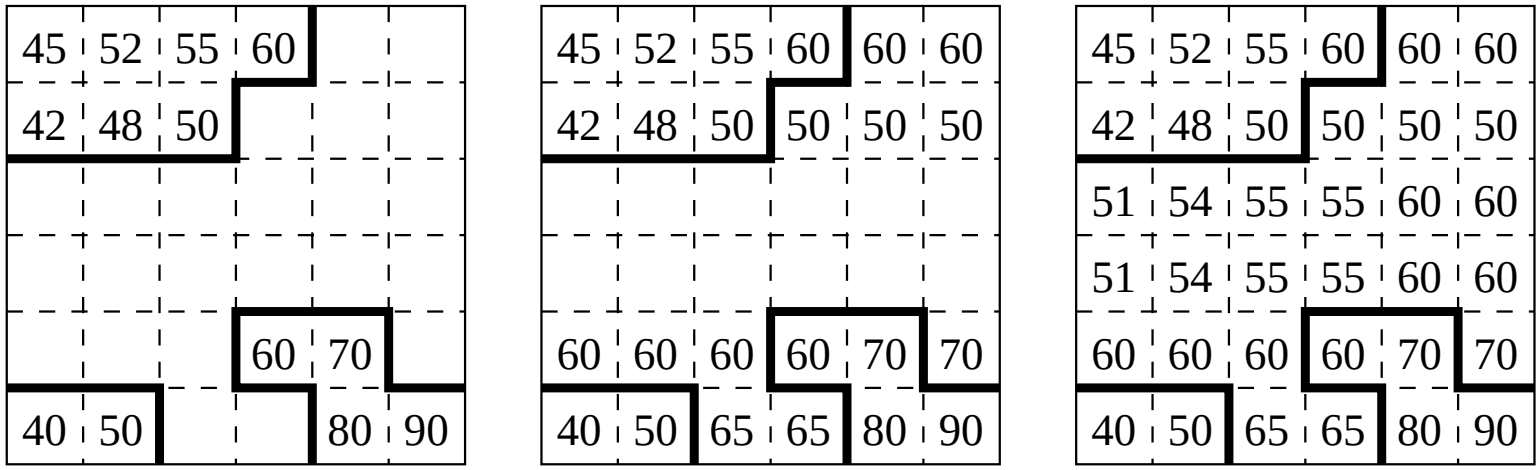

Fig. 12.7: An example of Repetitive Padding in a boundary macroblock of a Reference VOP: (a) Original pixels within the VOP, (b) After Horizontal Repetitive Padding, (c) Followed by Vertical Repetitive Padding.

II. Motion Vector Coding

- Let $C(x+k, y+l)$ be pixels of the MB in Target VOP, and $R(x+i+k, y+j+l)$ be pixels of the MB in Reference VOP.
- A **Sum of Absolute Difference (SAD)** for measuring the difference between the two MBs can be defined as:

$$SAD(i,j) = \sum_{k=0}^{N-1}\sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+i+k, y+j+l)| \cdot Map(x+k, y+l)$$

  N — the size of the MB. $Map(p,q) = 1$ when $C(p,q)$ is a pixel within the target VOP, otherwise $Map(p,q) = 0$.

- The vector (i,j) that yields the minimum SAD is adopted as the motion vector $\mathbf{MV}(u,v)$:

$$(u,v) = [\ (i,j) \mid SAD(i,j) \text{ is minimum, } i \in [-p,p],\ j \in [-p,p]\ ] \quad (12.1)$$

  p — the maximal allowable magnitude for u and v.

12.6 MPEG-7

- The main objective of MPEG-7 is to serve the need of audio-visual content-based retrieval (or audiovisual object retrieval) in applications such as digital libraries.
- Nevertheless, it is also applicable to any multimedia applications involving the generation (*content creation*) and usage (*content consumption*) of multimedia data.
- MPEG-7 became an International Standard in September 2001 — with the formal name **Multimedia Content Description Interface**.

Applications Supported by MPEG-7

- MPEG-7 supports a variety of multimedia applications. Its data may include still pictures, graphics, 3D models, audio, speech, video, and composition information (how to combine these elements).
- These MPEG-7 data elements can be represented in textual format, or binary format, or both.
- Fig. 12.17 illustrates some possible applications that will benefit from the MPEG-7 standard.

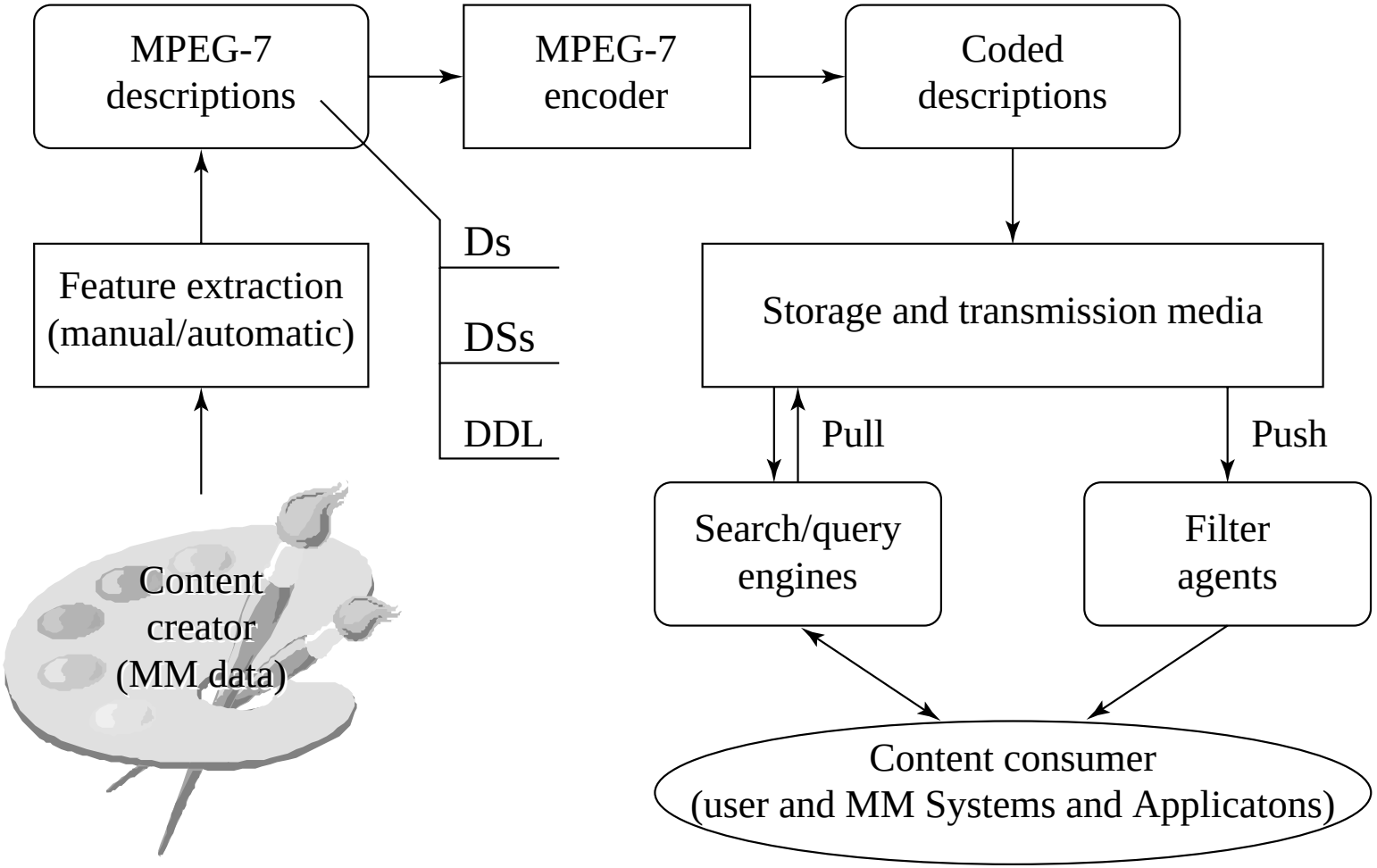

Fig. 12.17: Possible Applications using MPEG-7.

MPEG-7 and Multimedia Content Description

- MPEG-7 has developed Descriptors **(D)**, Description Schemes **(DS)** and Description Definition Language **(DDL)**. The following are some of the important terms:
  - **Feature** — characteristic of the data.
  - **Description** — a set of instantiated Ds and DSs that describes the structural and conceptual information of the content, the storage and usage of the content, etc.
  - **D** — definition (syntax and semantics) of the feature.
  - **DS** — specification of the structure and relationship between Ds and between DSs.
  - **DDL** — syntactic rules to express and combine DSs and Ds.
- The scope of MPEG-7 is to standardize the Ds, DSs and DDL for descriptions. The mechanism and process of producing and consuming the descriptions are beyond the scope of MPEG-7.

Descriptor (D)

- The descriptors are chosen based on a comparison of their performance, efficiency, and size. Low-level visual descriptors for basic visual features include:
  - **Color**
    - Color space. (a) RGB, (b) YCbCr, (c) HSV (hue, saturation, value), (d) HMMD (HueMaxMinDiff), (e) 3D color space derivable by a 3×3 matrix from RGB, (f) monochrome.
    - Color quantization. (a) Linear, (b) nonlinear, (c) lookup tables.
    - Dominant colors.
    - Scalable color.
    - Color layout.
    - Color structure.
    - Group of Frames/Group of Pictures (GoF/GoP) color.

- **Texture**
  - Homogeneous texture.
  - Texture browsing.
  - Edge histogram.
- **Shape**
  - Region-based shape.
  - Contour-based shape.
  - 3D shape.

- **Motion**
  - Camera motion (see Fig. 12.18).
  - Object motion trajectory.
  - Parametric object motion.
  - Motion activity.
- **Localization**
  - Region locator.
  - Spatiotemporal locator.
- **Others**
  - Face recognition.

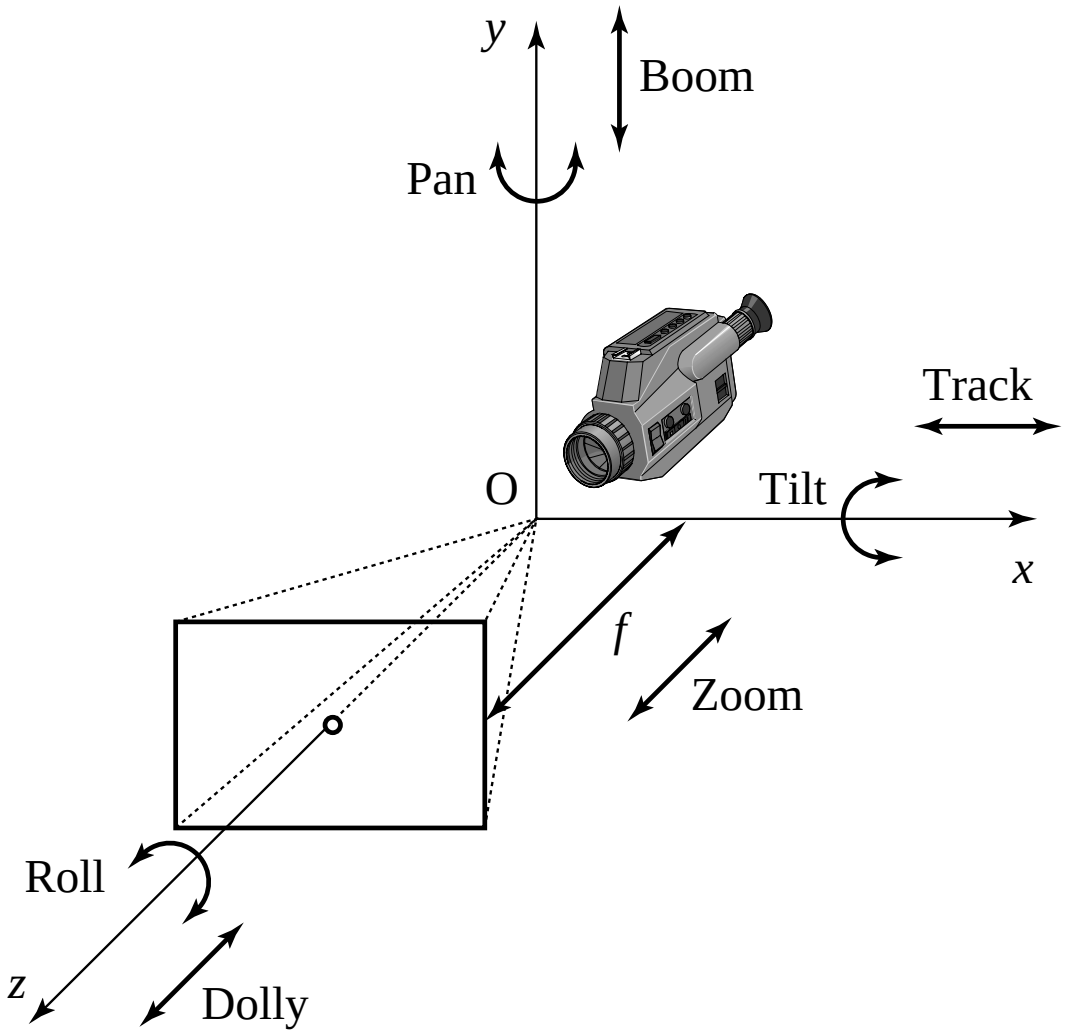

Fig. 12.18: Camera motions: pan, tilt, roll, dolly, track, and boom.

Description Scheme (DS)

- **Basic elements**
  - Datatypes and mathematical structures.
  - Constructs.
  - Schema tools.
- **Content Management**
  - Media Description.
  - Creation and Production Description.
  - Content Usage Description.
- **Content Description**
  - Structural Description.

A *Segment* DS, for example, can be implemented as a class object. It can have five subclasses: *Audiovisual segment DS*, *Audio segment DS*, *Still region DS*, *Moving region DS*, and *Video segment DS*. The subclass DSs can recursively have their own subclasses.

  - Conceptual Description.

- **Navigation and access**
  - Summaries.
  - Partitions and Decompositions.
  - Variations of the Content.
- **Content Organization**
  - Collections.
  - Models.
- **User Interaction**
  - UserPreference.

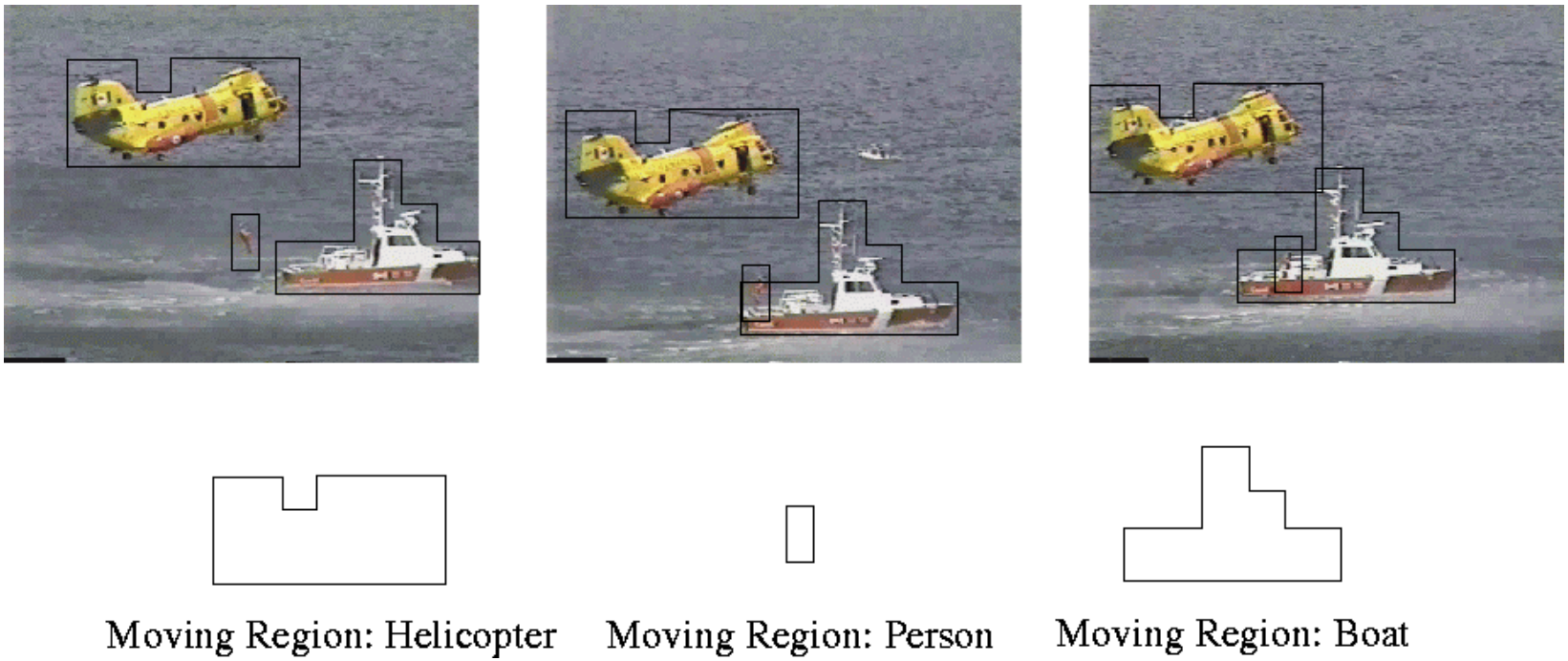

Fig. 12.19: MPEG-7 video segment.

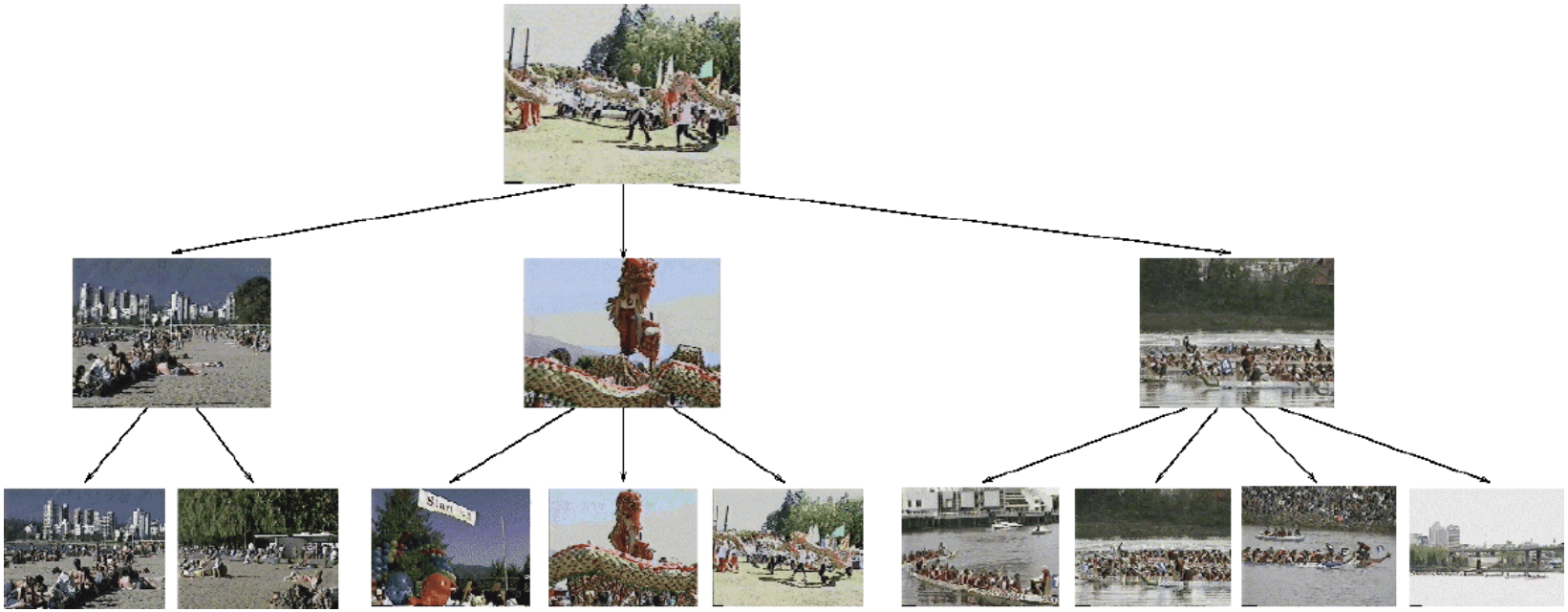

Fig. 12.20: A video summary.

Description Definition Language (DDL)

- MPEG-7 adopted the XML Schema Language initially developed by the WWW Consortium (W3C) as its Description Definition Language (DDL). Since XML Schema Language was not designed specifically for audiovisual contents, some extensions are made to it:
  - Array and matrix data types.
  - Multiple media types, including audio, video, and audiovisual presentations.
  - Enumerated data types for **MimeType, CountryCode, RegionCode, CurrencyCode,** and **CharacterSetCode**.
  - Intellectual Property Management and Protection **(IPMP)** for Ds and DSs.

12.7 MPEG-21

- The development of the newest standard, **MPEG-21: Multimedia Framework**, started in June 2000, and was expected to become International Stardard by 2003.
- The *vision* for MPEG-21 is to define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities.
- The seven key elements in MPEG-21 are:
  - **Digital item declaration** — to establish a uniform and flexible abstraction and interoperable schema for declaring Digital items.
  - **Digital item identification and description** — to establish a framework for standardized identification and description of digital items regardless of their origin, type or granularity.

- **Content management and usage** — to provide an interface and protocol that facilitate the management and usage (searching, caching, archiving, distributing, etc.) of the content.
- **Intellectual property management and protection (IPMP)** — to enable contents to be reliably managed and protected.
- **Terminals and networks** — to provide interoperable and transparent access to content with Quality of Service (QoS) across a wide range of networks and terminals.
- **Content representation** — to represent content in an adequate way for pursuing the objective of MPEG-21, namely "content anytime anywhere".
- **Event reporting** — to establish metrics and interfaces for reporting *events* (user interactions) so as to understand performance and alternatives.

12.8 Further Exploration

- **Text books:**
  - *Multimedia Systems, Standards, and Networks* by A. Puri and T. Chen
  - *The MPEG-4 Book* by F. Pereira and T. Ebrahimi
  - *Introduction to MPEG-7: Multimedia Content Description Interface* by B.S. Manjunath et al.
- **Web sites:** ⟶ Link to Further Exploration for Chapter 12.. including:
  - The MPEG home page
  - The MPEG FAQ page
  - Overviews, tutorials, and working documents of MPEG-4
  - Tutorials on MPEG-4 Part 10/H.264
  - Overviews of MPEG-7 and working documents for MPEG-21
  - Documentation for XML schemas that form the basis of MPEG-7 DDL