INTERNATIONAL STANDARD

ISO/IEC 19566-8

First edition

Information technologies — JPEG systems —

Part 8: JPEG Snack

Technologies de l'information — Systèmes JPEG —

Partie 8: Métadonnées d'enrichissement destinées à faciliter la consommation des contenus JPEG

PROOF/ÉPREUVE

Reference number
ISO/IEC 19566-8:2022(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives or www.iec.ch/members_experts/refdocs).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see https://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html. In the IEC, see www.iec.ch/understanding-standards.

This document was prepared by Joint Technical Committee ISO/JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

A list of all parts in the ISO/IEC 19566 series can be found on the ISO and IEC websites.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html and www.iec.ch/national-committees.

Introduction

The ISO/IEC 19566 series, on JPEG systems, contributes to the specification of system-level functionalities.

JPEG Snack is a means to convey relatively simple multimedia experiences which is fundamentally based on images and the image file format. Many digital storytelling experiences are based on converting images into video-based technologies, whereas images are directly used in JPEG Snack, along with playback of other media (video, audio, titles, captions, and effects) coordinated through an explicit timeline.

Information technologies — JPEG systems —

Part 8:
JPEG Snack

1 Scope

This document defines JPEG Snack metadata that enriches a representation of multiple media contents, in order to facilitate sharing, editing, and presentation; it further specifies metadata and container formats for JPEG Snack format.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 10918-1, *Information technology — Digital compression and coding of continuous-tone still images: Requirements and guidelines*

ISO/IEC 15444-2, *Information technology — Part 2: Extensions*

ISO/IEC 18477-3, *Information technology — Scalable compression and coding of continuous-tone still images — Part 3: Box file format*

ISO/IEC 19566-5, *Information technology — Part 5: JPEG Universal Metadata Box Format (JUMBF)*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 10918-1 and ISO/IEC 18477-3 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
snack culture
consumption of image-rich media in a short story format

3.2
media type
indicator of the format and content of the file transmitted through the Internet.

3.3
z-order
ordering of overlapping two dimensional regions that define the occlusion precedence amongst them

4 Overview

This document specifies metadata and formats that enable storing, sharing, and rendering snack culture contents with JPEG image coding standards.

NOTE The snack culture contents are defined as follows:

— image sequence from which one or more frames are generated by manipulating still images;

— image sequence recorded with a short playing duration, e.g. 1.5 s;

— image sequence with transition effects and/or overlay along with subtitles, audio clips, and graphics.

JPEG Snack is a format that defines the representation of multimedia, such as images, image sequences, text, audio, and video clips, including transition effects, based on the existing JPEG family image coding standards. Besides, it supports a timing mechanism to synchronize multimedia with a global timeline in a context. This mechanism allows users to watch multimedia contents like short-form video clips. However, unlike conventional video formats, it supports storing images without transcoding from image to dedicated video codec.

In order to define the functionalities of the JPEG Snack format, this document is organized as follows:

— 4.1 describes the overall system of the JPEG Snack format.

— 4.2 describes the system decoder model.

— 4.3 defines an essential model of metadata to compose the JPEG Snack format.

— Clauses 5 and 6 describe the JPEG Snack format in detail.

— Annexes A to C explain how the metadata is serialized and describe the formation of the JPEG Snack file and its usage examples.

4.1 System description

This document specifies metadata and its behaviour to compose the JPEG Snack content by synchronizing multimedia on the decoder side. This document primarily defines a metadata model consisting of two formats:

— Object-structured format: describes the content and additional behaviours of the objects are structured in the object-composition description.

— Object-composition format: describes the positional and temporal relationships between objects and the composition of the objects onto the decoder display.

Its hierarchical structure of the JPEG Snack format is depicted in Figure 4.1.

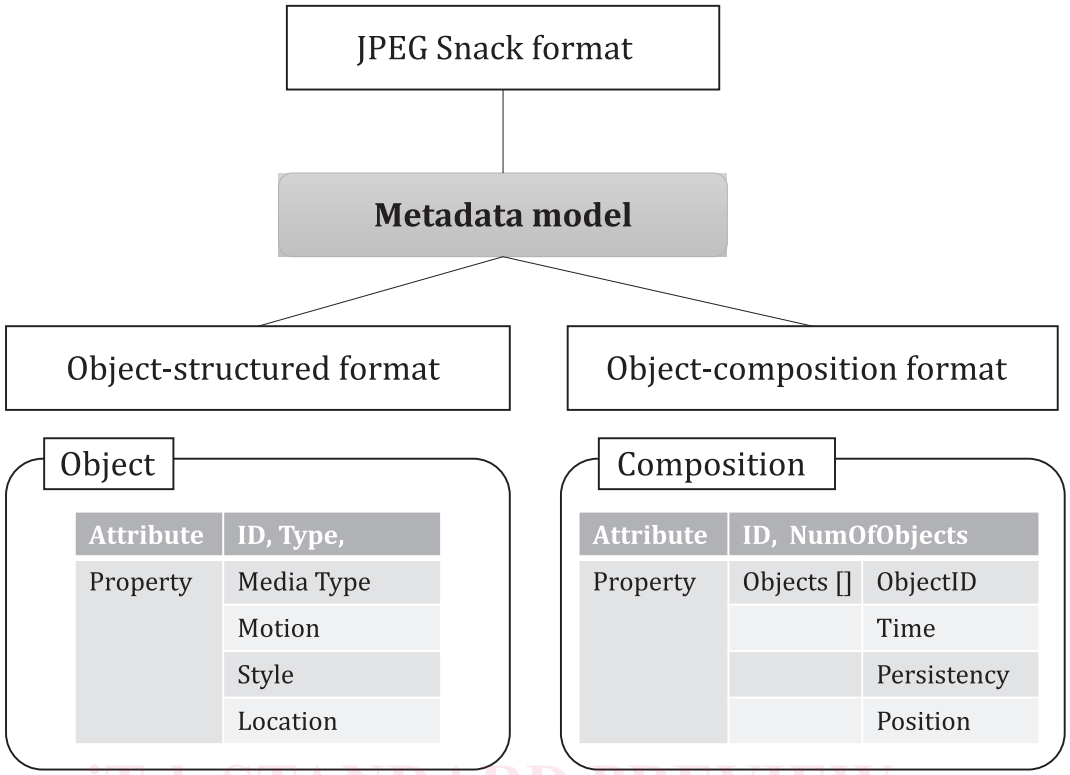


Figure 4.1 — Overview of the JPEG Snack format

The JPEG Snack format provides information that enables JPEG Snack applications to share and render media contents by accessing the objects in the file or reference to objects contained in other files. All objects are not necessarily embedded in the same file. Each object constituting a JPEG Snack file is structured using a box defined in ISO/IEC 19566 and stored into a JPEG image file.

The object-structured format defines the appearance and behaviour of the individual object. This format includes the size and opacity of the object, movement information in a given timeline of the representation, and information on the location where the media data, such as an image codestream, is found (see Clause 5).

The object-composition format identifies the objects that compose the representation and defines each object's creation and destruction. This format describes the temporal and spatial relationship between objects by providing information on the time and position of the individual object to show, and the time and position of their disappearance. Each object has independent position information on the decoder screen, and the composition information determines the z-order of the objects displayed to the user (see Clause 6).

4.2 System decoder model

A JPEG Snack decoder implements the metadata model described in 4.1. The decoder has three conceptual necessary components: default image, timeline, and layer and position, as depicted in Figure 4.2. The decoder decodes the JPEG image to prepare a default image and compose a JPEG Snack representation with several objects using this default image as a background. Since the JPEG Snack is created by defining when, where and how objects are composed, the decoder shall handle timeline, layer, and position.

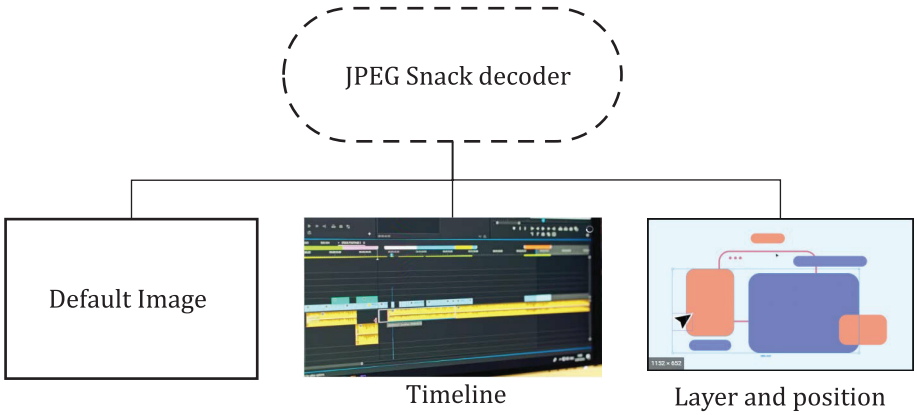
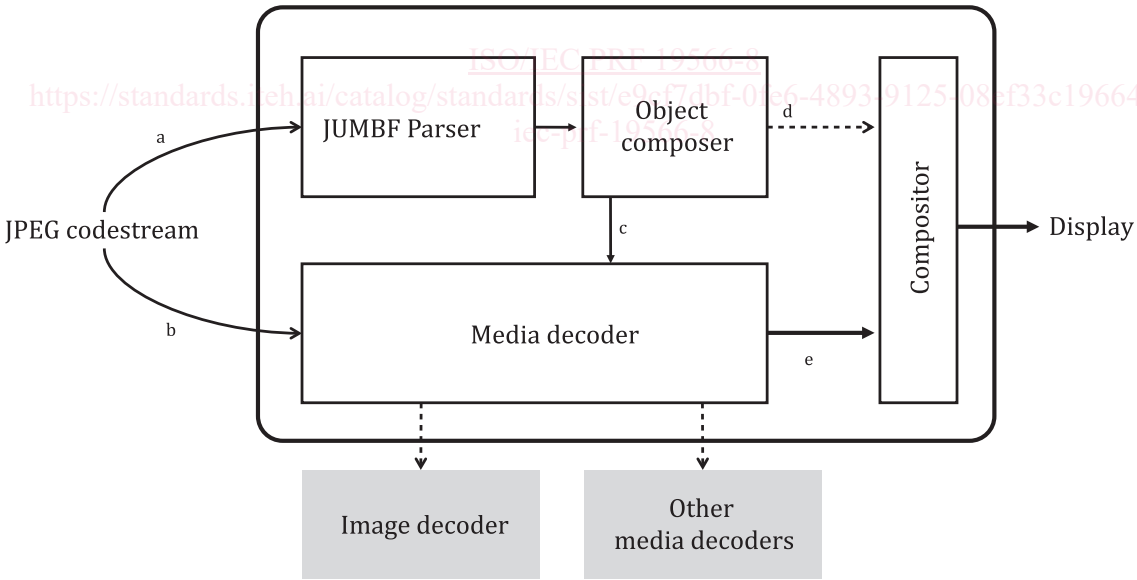


Figure 4.2 — Overview of the JPEG Snack decoder

This document defines the formats based on the informative system decoder model of JPEG Snack, as depicted in Figure 4.3, to allow various JPEG image coding standards to represent JPEG Snack contents in a concerted way. Figure 4.3 illustrates an example of the JPEG Snack decoder in which the formats defined in 4.1 may be implemented.

In Figure 4.3, the object composer receives a JPEG codestream that contains metadata and media data through the JUMBF parser, constructs the JPEG Snack representation, invokes media decoders to decode its media data from the codestream, and renders the media content decoded to the output devices. The object composer controls the media decoder and compositor to decode and display its media content regarding time and position appropriately. This version of the document allows images, captions, image sequences, audio clips, video clips to be composed in a representation of JPEG Snack.



a Metadata.
b Media data.
c Media format + time.
d Position + z-order.
e Media output.

Figure 4.3 — Overview of the system decoder model for JPEG Snack

4.3 Metadata model

The system decoder model described in 4.2 is based on the JPEG Snack format depicted in Figure 4.1 to support the playback of JPEG Snack contents being constituted by multiple media contents.

The metadata is a hierarchical model, as illustrated in Figure 4.4, containing multiple object metadata (see Clause 6) aligned with composition metadata corresponded to the object-composition format. Within the object metadata corresponded to the object-structured format, properties (see Annex A) composing the objects into a representation of the JPEG Snack format such as position, time, and transition are contained. Each object may be rendered individually in a logical timeline of the decoder to support re-editing the object; for example, a user may choose a specific object to hide in his/her JPEG Snack viewer.

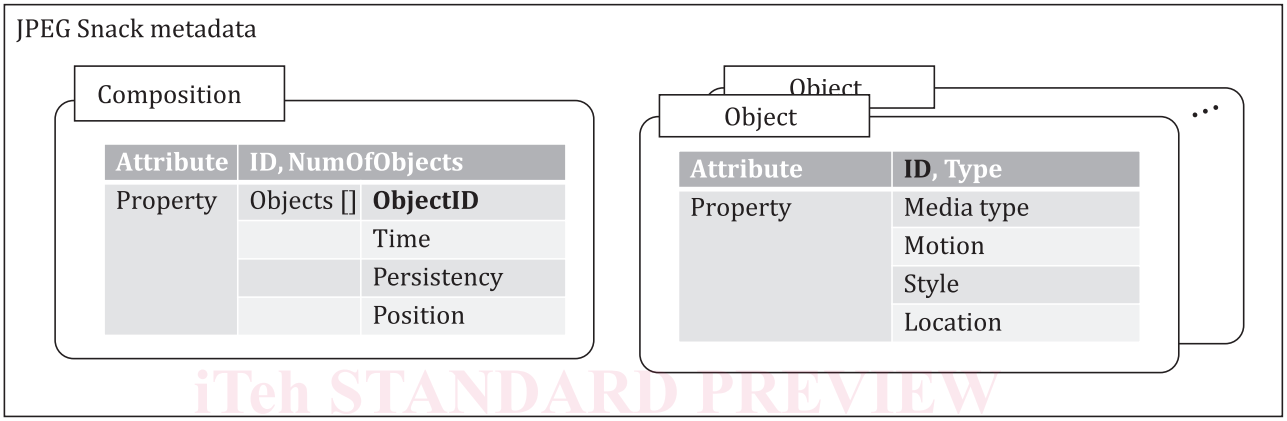


Figure 4.4 — High-level metadata model of JPEG Snack

Object metadata specifies the content and additional behaviour of the individual objects that compose the representation and identifies where the object's resides. An ID is an identifier of the object in the representation and a Type attribute allows a decoder to recognize properties of the object proactively.

Composition metadata coordinates the objects composing a JPEG Snack representation. The objects are arranged into Objects within a composition along with position and time with an identifier attribute. A Position property determines where the object pointed to by the ObjectID is placed. When objects are overlapped according to the Position property, the Time and Persistency properties organize the objects to be placed in front or behind the other object (see 6.2).

JPEG Snack shall have only one composition metadata consisting of one or more objects within a scope of the JPEG Snack file.

The JPEG Snack decoder described in 4.2 composes a timeline (see 6.1.2) for playback of the JPEG Snack content by combining the Time information of all objects, and they exist in the representation individually using their Position and Time information.

4.4 Object-structured file organization

An object in the file organization is a JUMBF box. The JPEG Snack files are formed as a series of boxes. All metadata is contained in boxes, as illustrated in Figure 4.5. JUMBF boxes for JPEG Snack contains metadata to compose the JPEG Snack representation, and other types of JUMBF box are used to deliver the media content, such as a codestream and XML document for each object. The boxes shall be embedded as defined in Annex A and ISO/IEC 19566-5.

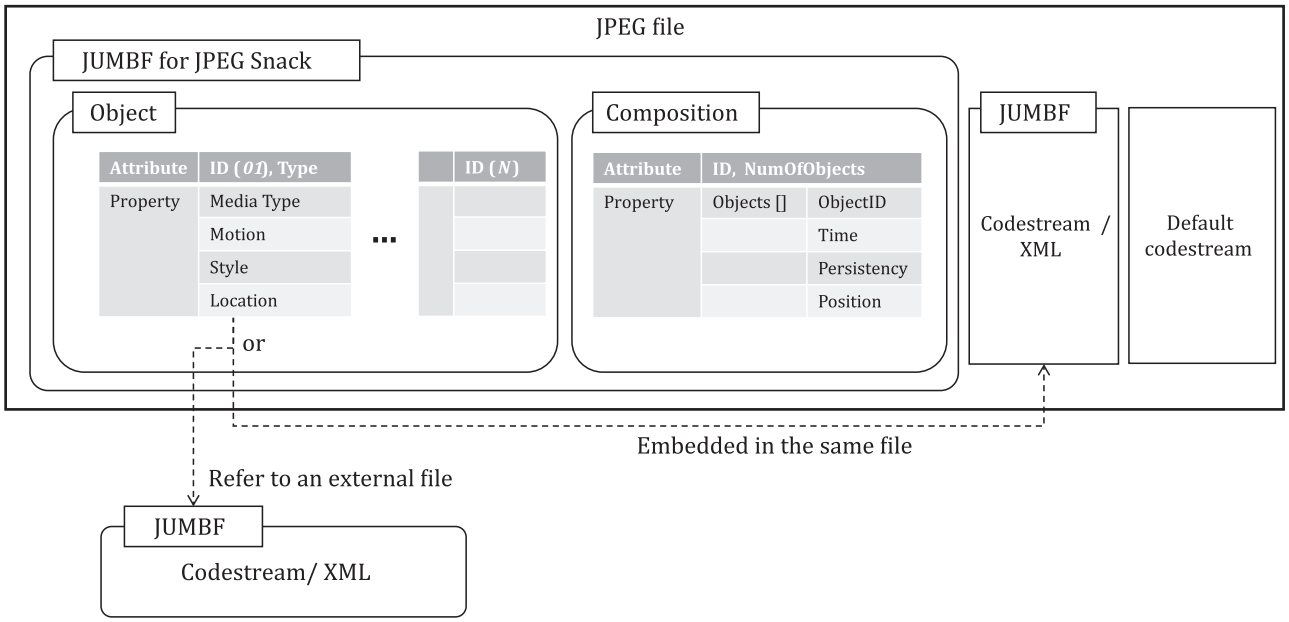


Figure 4.5 — Organization of the JPEG Snack file

The JPEG Snack format provides information to define the metadata for composing the representation and the format in which the metadata is structured in the JPEG image files. The JPEG Snack file has a different file extension according to the default codestream. Conventional JPEG decoders may ignore JUMBF boxes for the JPEG. For example, if the JPEG Snack metadata is embedded in the file of the ISO/IEC 10918-1, denoted by JPEG-1, the extension of the JPEG Snack file is '.jpg' like conventional JPEG-1 images while the conventional JPEG-1 decoder decodes only the default codestream. This feature provides compatibility to the existing JPEG image coding standards, including future standards based on the box-based format.

NOTE 1 The default codestream is placed at the end of the file to be compatible with the conventional JPEG image coding standards. For example, the JPEG-1 decoder can ignore any extra data beyond the EOI (end of image) marker.

NOTE 2 Codestream is a sequence of bits representing a compressed image and associated metadata.

In addition, content types of which is indicated by the object metadata may be different JUMBF boxes based on the object type. The object may refer to JUMBF boxes for media data embedded in another file. The referencing shall be done as defined in ISO/IEC 19566-5:2019, Annex C.

5 Object-structured format

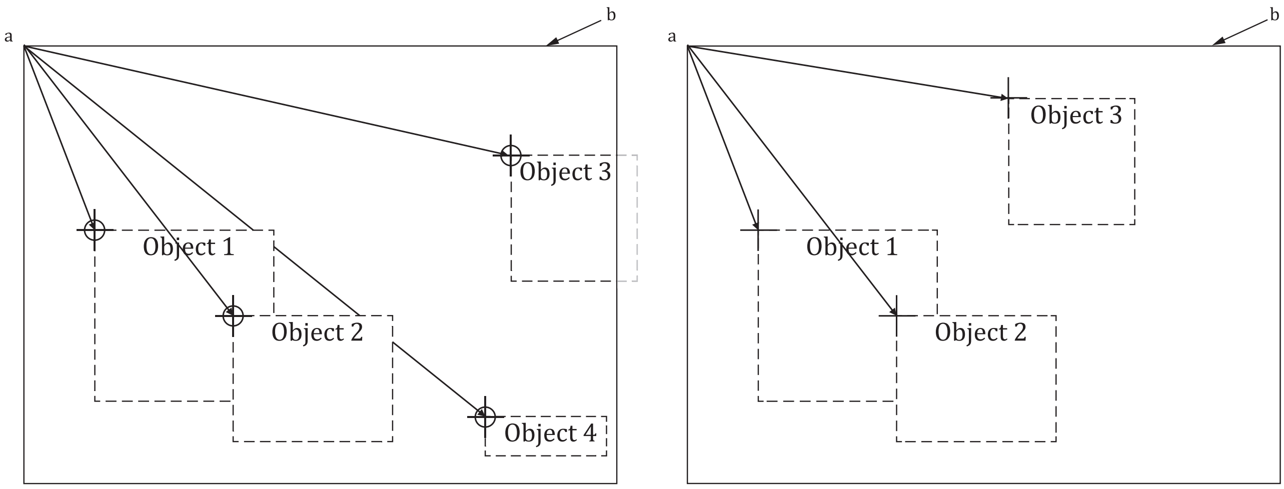
5.1 General

As described in Clause 4, in the JPEG Snack format, the representation of the JPEG Snack is composed of a group of media contents. The object in this document is a unit that composes a JPEG Snack format and contains information to represent the media contents.

Figures 5.1 and 5.2 illustrate the roles of the object-composition and object-structured formats to compose JPEG Snack representation. The object-composition format (see Clause 6) provides composition information to define when and where the objects that are constructed will appear and disappear in a representation, whereas the object-structured format signals information on the individual object's behavior and location of the resource. In Figure 4.3, while the object composer manages instances of the object, the decoding of the individual object is conducted independently by the media decoder. The

object composer informs the compositor z-order and movement information of the object. Then the compositor renders the decoded media data accordingly based on the z-order and position information.

NOTE An invisible object, such as an audio clip, does not have z-order and position information. And a description of spatial audio is not included in this document, whereas it is considered as a typical audio clip.



a) Composition of objects at t_0 **b) Composition of objects at t_1**

Key

t_0 time when the representation is started

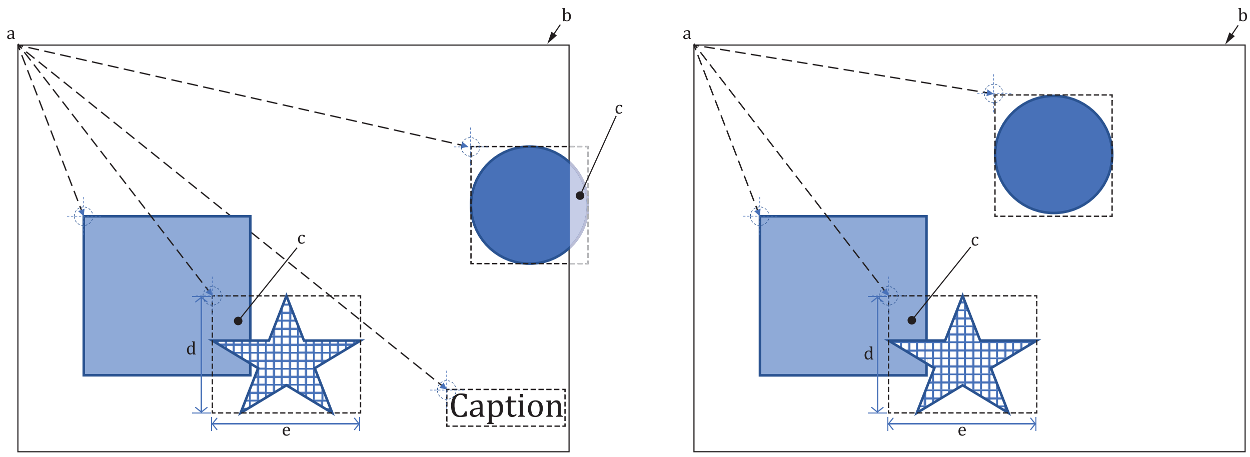
t_1 time when the representation is ended

a Origin.

b Representation.

Figure 5.1 — Example of the object-composition format. The t_0 is a time when the representation is started and the t_1 is when the representation is ended.

In the example of Figure 5.1, object 2 is above object 1 so that object 1 has an occluded region. Also, object 3 has an occluded region beyond a representation. The object composer shall handle these regions smoothly. For object 4, the duration of existence is shorter than the JPEG Snack's total duration. See Clause 6.1.2 for more details on the temporal composition of objects.



a) Objects' shape at t_0 **b) Objects' shape at t_1**