INTERNATIONAL STANDARD

ISO/IEC 23090-7

First edition
2022-11

Information technology — Coded representation of immersive media —

Part 7: Immersive media metadata

Technologies de l'information — Représentation codée de média immersifs —
Partie 7: Métadonnées de media immersifs

Reference number
ISO/IEC 23090-7:2022(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/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

A list of all parts in the ISO/IEC 23090 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

This document is organized as follows.

— Clauses 5 describes common metadata applicable to immersive media. This includes reference coordinate system related metadata and other common metadata syntax and semantics.

— Clauses 6 describes metadata that applies to video and images. This includes projection formats and packing region-wise formats metadata which applies to video and images.

The goal of this document is to allow reuse of the commonly defined metadata to be referenced by other standards.

The International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) draw attention to the fact that it is claimed that compliance with this document may involve the use of a patent.

ISO and IEC take no position concerning the evidence, validity and scope of this patent right.

The holder of this patent right has assured ISO and IEC that he/she is willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statement of the holder of this patent right is registered with ISO and IEC. Information may be obtained from the patent database available at www.iso.org/patents or https://patents.iec.ch.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights other than those in the patent database. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

Information technology — Coded representation of immersive media —

Part 7: Immersive media metadata

1 Scope

This document specifies common immersive media metadata focusing on immersive videos (including 360° videos) and images.

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 14496-12, *Information technology — Coding of audio-visual objects — Part 12: ISO base media file format*

ISO/IEC 23008-12, *Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 12: Image file format*

3 Terms, definitions and symbols

3.1 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 14496-12 and ISO/IEC 23008-12 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/

3.1.1
azimuth
first of the two *sphere coordinates* (3.1.22) describing the location of a point on the sphere

3.1.2
azimuth circle
circle on the sphere connecting all points with the same *azimuth* (3.1.1) value

Note 1 to entry: An azimuth circle is always a *great circle* (3.1.12).

3.1.3
circular image
image captured with a *fisheye lens* (3.1.9)

3.1.4
common reference coordinate system
3D Cartesian coordinate system with the centre being (X, Y, Z) equal to (0, 0, 0), used as the reference coordinate system for all viewpoints within a *viewpoint group* (3.1.27)

3.1.5
content coverage
one or more *sphere regions* (3.1.23) that are covered by the content represented by the track or by an image item

3.1.6
elevation
second of the two *sphere coordinates* (3.1.22) describing the location of a point on the sphere

3.1.7
elevation circle
circle on the sphere connecting all points with the same *elevation* (3.1.6) value

Note 1 to entry: When the elevation is zero, an elevation circle is also a *great circle* (3.1.12). This coincides with the equator on Earth.

3.1.8
field of view
extent of the observable world in captured/recorded content or in a physical display device

3.1.9
fisheye lens
wide-angle camera lens that usually captures an approximately hemispherical *field of view* (3.1.8) and projects it as a *circular image* (3.1.3)

3.1.10
fisheye video
video captured by *fisheye lenses* (3.1.9)

3.1.11
global coordinate axes
coordinate axes that are associated with audio, video, and images representing the same acquisition position and intended to be rendered together

3.1.12
great circle
intersection of the sphere and a plane that passes through the centre point of the sphere

Note 1 to entry: A great circle is also known as an orthodrome or Riemannian circle.

Note 2 to entry: The centre of the sphere and the centre of a great circle are co-located.

3.1.13
guard band
area in a *packed picture* (3.1.16) that is not rendered but may be used to improve the rendered part of the packed picture to avoid or mitigate visual artifacts such as seams

Note 1 to entry: Guard bands are associated with *packed regions* (3.1.17) as described in 6.5.2.

3.1.14
local coordinate axes
coordinate axes obtained after applying rotation to the *global coordinate axes* (3.1.11)

3.1.15
omnidirectional video
video and its associated audio that enable rendering according to the user's *viewing orientation* (3.1.26), if consumed with a head-mounted device, or according to user's desired *viewport* (3.1.28), otherwise, as if the user was in the spot where and when the media was captured

3.1.16
packed picture
picture that is represented as a coded picture in the coded video bitstream

3.1.17
packed region
region in a *packed picture* (3.1.16) that is mapped to a *projected region* (3.1.19) as specified by the *region-wise packing* (3.1.21) signalling

3.1.18
projected picture
picture that has a representation format specified by an *omnidirectional video* (3.1.15) *projection* (3.1.20) format

3.1.19
projected region
region in a *projected picture* (3.1.18) that is mapped to a *packed region* (3.1.17) as specified by the *region-wise packing* (3.1.21) signalling

3.1.20
projection
inverse of the process by which the samples of a *projected picture* (3.1.18) are mapped to a set of positions identified by a set of *azimuth* (3.1.1) and *elevation* (3.1.6) coordinates on a unit sphere

3.1.21
region-wise packing

inverse of the process of transformation, resizing, and relocating of *packed regions* (3.1.17) of a *packed picture* (3.1.16) to remap to *projected regions* (3.1.19) of a *projected picture* (3.1.18)

3.1.22
sphere coordinates
azimuth (ϕ) (3.1.1) and *elevation* (θ) (3.1.6) that identify a location of a point on the unit sphere

3.1.23
sphere region
region on a sphere, specified either by four *great circles* (3.1.12) or by two *azimuth circles* (3.1.2) and two *elevation circles* (3.1.7), or such a region on the rotated sphere after applying certain amount of yaw, pitch, and roll rotations

3.1.24
SDL
syntactic description language
language that allows the description of a bitstream's syntax

Note 1 to entry: Syntactic description language is defined in ISO/IEC 14496-1:2010, Clause 8.

3.1.25
tilt angle
angle indicating the amount of tilt of a *sphere region* (3.1.23), measured as the amount of rotation of the sphere region along the axis originating from the sphere origin passing through the centre point of the sphere region, where the angle value increases clockwise when looking from the origin towards the positive end of the axis

3.1.26
viewing orientation
triple of *azimuth* (3.1.1), *elevation* (3.1.6), and *tilt angle* (3.1.25) characterizing the orientation that a user is consuming the audio-visual content

Note 1 to entry: In case of image or video, viewing orientation characterizes the orientation of the *viewport* (3.1.28).

3.1.27
viewpoint group
group of viewpoints that share the same *common reference coordinate system* (3.1.4)

3.1.28
viewport
region of omnidirectional image or video suitable for display and viewing by the user

3.2 Symbols

+	Addition.
−	Subtraction (as a two-argument operator) or negation (as a unary prefix operator).
*	Multiplication, including matrix multiplication.
x^y	Exponentiation. Specifies *x* to the power of *y*. In other contexts, such notation is used for superscripting not intended for interpretation as exponentiation.
/	Integer division with truncation of the result toward zero. For example, 7 / 4 and −7 / −4 are truncated to 1 and −7 / 4 and 7 / −4 are truncated to −1.
÷	Used to denote division in mathematical equations where no truncation or rounding is intended.
$\frac{x}{y}$	Used to denote division in mathematical equations where no truncation or rounding is intended.
$\sum_{i=x}^{y} f(i)$	The summation of f(*i*) with *i* taking all integer values from *x* up to and including *y*.
x % *y*	Modulus. Remainder of *x* divided by *y*, defined only for integers *x* and *y* with *x* >= 0 and y > 0.
Asin(*x*)	The trigonometric inverse sine function, operating on an argument *x* that is in the range of −1.0 to 1.0, inclusive, with an output value in the range of −π÷2 to π÷2, inclusive, in units of radians.
Atan(*x*)	The trigonometric invers tangent function, operating on an argument *x* that is any real number, with an output value in the range of −π÷2 to π÷2, inclusive, in units of radians.

$$\mathrm{Atan2}(y,x)=\begin{cases}\mathrm{Atan}\left(\frac{y}{x}\right) & ;\ \text{if } x>0\\ \mathrm{Atan}\left(\frac{y}{x}\right)+\pi & ;\ \text{if } x<0\ \&\&\ y>=0\\ \mathrm{Atan}\left(\frac{y}{x}\right)-\pi & ;\ \text{if } x<0\ \&\&\ y<0\\ +\frac{\pi}{2} & ;\ \text{if } x==0\ \&\&\ y>=0\\ -\frac{\pi}{2} & ;\ \text{otherwise}\end{cases} \tag{3-1}$$

Cos(*x*) The trigonometric cosine function operating on an argument *x* in units of radians.

Floor(*x*) The the largest integer less than or equal to *x*.

Sin(*x*) The trigonometric sine function operating on an argument *x* in units of radians.

Tan(*x*) The trigonometric tangent function operating on an argument *x* in units of radians.

4 Overview

4.1 General

This document specifies common immersive media metadata focusing on immersive videos (including 360° videos) and images. The metadata includes co-ordinate system, projection format, and packing region-wise formats metadata.

4.2 Variables

This document derives variables that are named by a mixture of lower case and upper case letter and without any underscore characters.

4.3 Processes

Processes are used to describe the various operations. A process has a set of one or more inputs, a set of one or more outputs and a sequence of operation steps.

4.4 Syntax structures

Syntax structures in this document are specified with the syntactic description language (SDL) specified in ISO/IEC 14496-1:2010, Clause 8, with the following change: Unlike specified in ISO/IEC 14496-1:2010, Clause 8, this document allows a variable declaration in `expression1` of a for loop `for(expression1; expression2; expression3)`. Such a variable declaration may be used for a loop index variable with a data type.

NOTE As specified in ISO/IEC 14496-1:2010, 8.3.6, this document allows declaring a syntax element that is an individual element in an array. Such a declaration follows ISO/IEC 14496-1:2010, Rule A.2: `typespec name[[index]]`; which declares the `index`-th element of the array `name` as an individual syntax element having the data `typespec`. In the context of this document, `typespec name[[index]]` is only used to refer to the index in the semantics and is actually equivalent to `typespec name`.

5 Common metadata

5.1 Reference coordinate system

The coordinate system consists of a unit sphere and three coordinate axes, namely the X (back-to-front) axis, the Y (lateral, side-to-side) axis, and the Z (vertical, up) axis, where the three axes cross at the centre of the sphere.

The location of a point on the sphere is identified by a pair of sphere coordinates azimuth (ϕ) and elevation (θ).

Figure 5.1 specifies the relation of the sphere coordinates azimuth (ϕ) and elevation (θ) to the X, Y, and Z coordinate axes.

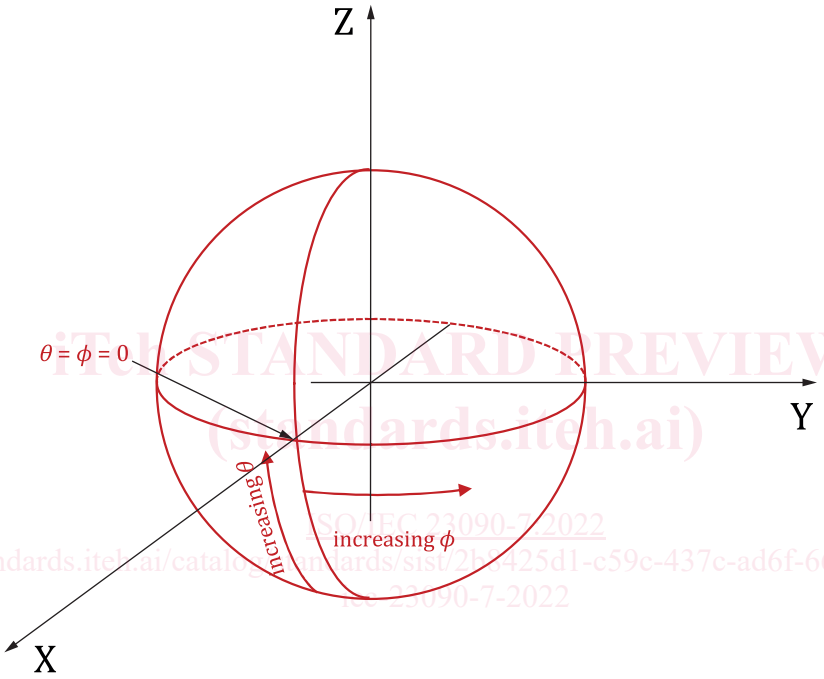

Figure 5.1 — Coordinate axes and their relation to the sphere coordinates

The value ranges of azimuth is −180.0, inclusive, to 180.0, exclusive, degrees. The value range of elevation is −90.0 to 90.0, inclusive, degrees.

5.2 Coordinate system rotation

Inputs to this process are:

— `rotation_yaw` (α_d), `rotation_pitch` (β_d), `rotation_roll` (γ_d), all in units of degrees, where `rotation_yaw` (α_d) and `rotation_roll` (γ_d), are in the range of −180.0, inclusive, to 180.0, exclusive, and `rotation_pitch` (β_d) is in the range of −90.0 to 90.0, inclusive, and

— sphere coordinates (ϕ_d, θ_d) relative to the local coordinate axes.

Outputs of this process are:

— sphere coordinates (ϕ', θ') in degrees relative to the global coordinate axes.

This process specifies rotations around the three axes of the coordinate system of 5.1 where yaw (α_d) expresses a rotation around the Z axis, pitch (β_d) rotates around the Y axis, and roll (γ_d) rotates around the X axis. Rotations are extrinsic, i.e. around X, Y, and Z fixed reference axes. The angles increase clockwise when looking from the origin towards the positive end of an axis, as illustrated in Figure 5.2.

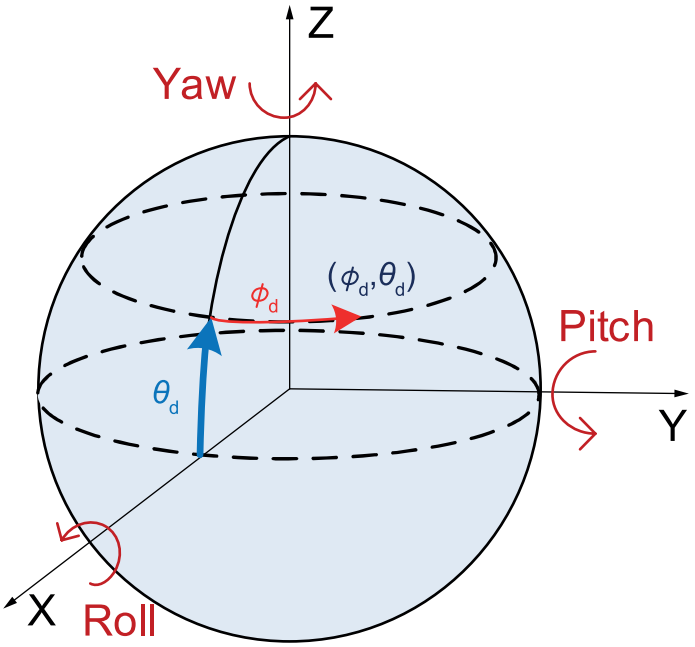

Figure 5.2 — Illustration of the directions of the yaw, pitch, and roll rotations

When any of the yaw (α_d), pitch (β_d) and roll (γ_d) rotation angles is not equal to zero, an OMAF player needs to apply the sphere rotation process specified in this clause to convert the local coordinate axes to the global coordinate axes.

It is assumed that the global coordinate systems for different media types were made aligned during content production.

The outputs are derived as follows:

$$\phi = \phi_d * \pi \div 180$$

$$\theta = \theta_d * \pi \div 180$$

$$\alpha = \alpha_d * \pi \div 180$$

$$\beta = \beta_d * \pi \div 180$$

$$\gamma = \gamma_d * \pi \div 180$$

$$x_1 = \mathrm{Cos}(\phi) * \mathrm{Cos}(\theta)$$

$$y_1 = \mathrm{Sin}(\phi) * \mathrm{Cos}(\theta)$$

$$z_1 = \mathrm{Sin}(\theta)$$

$$x_2 = \mathrm{Cos}(\beta) * \mathrm{Cos}(\alpha) * x_1 - \mathrm{Cos}(\beta) * \mathrm{Sin}(\alpha) * y_1 + \mathrm{Sin}(\beta) * z_1$$

$$y_2 = (\mathrm{Cos}(\gamma) * \mathrm{Sin}(\alpha) + \mathrm{Sin}(\gamma) * \mathrm{Sin}(\beta) * \mathrm{Cos}(\alpha)) * x_1 + (\mathrm{Cos}(\gamma) * \mathrm{Cos}(\alpha) - \mathrm{Sin}(\gamma) * \mathrm{Sin}(\beta) * \mathrm{Sin}(\alpha)) * y_1 - \mathrm{Sin}(\gamma) * \mathrm{Cos}(\beta) * z_1$$

$$z_2 = (\mathrm{Sin}(\gamma) * \mathrm{Sin}(\alpha) - \mathrm{Cos}(\gamma) * \mathrm{Sin}(\beta) * \mathrm{Cos}(\alpha)) * x_1 + (\mathrm{Sin}(\gamma) * \mathrm{Cos}(\alpha) + \mathrm{Cos}(\gamma) * \mathrm{Sin}(\beta) * \mathrm{Sin}(\alpha)) * y_1 + \mathrm{Cos}(\gamma) * \mathrm{Cos}(\beta) * z_1$$

$\phi' = \text{Atan2}(y_2, x_2) * 180 \div \pi$

$\theta' = \text{Asin}(z_2) * 180 \div \pi$

5.3 Common metadata data structures

5.3.1 Rotation structure

5.3.1.1 Definition

The fields in this structure provides the yaw, pitch, and roll angles, respectively, of the rotation to be applied to convert the local coordinate axes to the global coordinate axes. In the case of stereoscopic omnidirectional video, the fields apply to each view individually.

5.3.1.2 Syntax

```
aligned(8) class RotationStruct() {
   signed int(32) rotation_yaw;
   signed int(32) rotation_pitch;
   signed int(32) rotation_roll;
}
```

5.3.2 Content coverage structure

5.3.2.1 Definition

The fields in this structure provides the content coverage, which is expressed by one or more sphere regions covered by the content, relative to the global coordinate axes.

5.3.2.2 Syntax


```
aligned(8) class ContentCoverageStruct() {
   unsigned int(8) coverage_shape_type;
   unsigned int(8) num_regions;
   unsigned int(1) view_idc_presence_flag;
   if (view_idc_presence_flag == 0) {
      unsigned int(2) default_view_idc;
      bit(5) reserved = 0;
   } else
      bit(7) reserved = 0;
   for ( i = 0; i < num_regions; i++) {
      if (view_idc_presence_flag == 1) {
         unsigned int(2) view_idc[i];
         bit(6) reserved = 0;
      }
      SphereRegionStruct(1, 1);
   }
}
```

5.3.3 Viewpoint information structures

5.3.3.1 Definition

The `ViewpointPosStruct()`, `ViewpointGpsPositionStruct()`, `ViewpointGeomagneticInfoStruct()`, `ViewpointGlobalCoordinateSysRotationStruct()`, and `ViewpointGroupStruct()` provide information of a viewpoint, including (X, Y, Z) position of the viewpoint, GPS position of the viewpoint, geomagnetic position information for the viewpoint, and the yaw, pitch, and roll rotation angles of X, Y, and Z axes, respectively, of the global coordinate system of the viewpoint relative to the common reference coordinate system, and viewpoint group information.

5.3.3.2 Syntax

```
aligned(8) ViewpointPosStruct() {
   signed int(32) viewpoint_pos_x;
   signed int(32) viewpoint_pos_y;
   signed int(32) viewpoint_pos_z;
}
aligned(8) class ViewpointGpsPositionStruct() {
   signed int(32) viewpoint_gpspos_longitude;
   signed int(32) viewpoint_gpspos_latitude;
   signed int(32) viewpoint_gpspos_altitude;
}
aligned(8) class ViewpointGeomagneticInfoStruct() {
   signed int(32) viewpoint_geomagnetic_yaw;
   signed int(32) viewpoint_geomagnetic_pitch;
   signed int(32) viewpoint_geomagnetic_roll;
}
aligned(8) class ViewpointGlobalCoordinateSysRotationStruct() {
   signed int(32) viewpoint_gcs_yaw;
   signed int(32) viewpoint_gcs_pitch;
   signed int(32) viewpoint_gcs_roll;
}
aligned(8) class ViewpointGroupStruct() {
   unsigned int(8) vwpt_group_id;
   utf8string vwpt_group_description;
}
```

5.3.4 Sphere region structure

5.3.4.1 Definition

The sphere region structure (`SphereRegionStruct`) specifies a sphere region.

When `centre_tilt` is equal to 0, the sphere region specified by this structure is derived as follows:

— If both `azimuth_range` and `elevation_range` are equal to 0, the sphere region specified by this structure is a point on a spherical surface.

— Otherwise, the sphere region is defined using variables centreAzimuth, centreElevation, cAzimuth1, cAzimuth, cElevation1, and cElevation2 derived as follows:

$$
\begin{aligned}
&\text{centreAzimuth} = \texttt{centre\_azimuth} \div 65536 \\
&\text{centreElevation} = \texttt{centre\_elevation} \div 65536 \\
&\text{cAzimuth1} = (\texttt{centre\_azimuth} - \texttt{azimuth\_range} \div 2) \div 65536 \\
&\text{cAzimuth2} = (\texttt{centre\_azimuth} + \texttt{azimuth\_range} \div 2) \div 65536 \\
&\text{cElevation1} = (\texttt{centre\_elevation} - \texttt{elevation\_range} \div 2) \div 65536 \\
&\text{cElevation2} = (\texttt{centre\_elevation} + \texttt{elevation\_range} \div 2) \div 65536
\end{aligned}
\tag{5-1}
$$

The sphere region is defined as follows with reference to the shape type value specified in the semantics of the structure containing this instance of `SphereRegionStruct`:

— When the shape type value is equal to 0, the sphere region is specified by four great circles defined by four points cAzimuth1, cAzimuth2, cElevation1, cElevation2 and the centre point defined by centreAzimuth and centreElevation and as shown in Figure 5.3.

— When the shape type value is equal to 1, the sphere region is specified by two azimuth circles and two elevation circles defined by four points cAzimuth1, cAzimuth2, cElevation1, cElevation2 and the centre point defined by centreAzimuth and centreElevation and as shown in Figure 5.4.

When `centre_tilt` is not equal to 0, the sphere region is firstly derived as above and then a tilt rotation is applied along the axis originating from the sphere origin passing through the centre point of the sphere region, where the angle value increases clockwise when looking from the origin towards the positive end of the axis. The final sphere region is the one after applying the tilt rotation.