VIDEO IMAGE DECODING AND COMPOSING METHOD AND VIDEO IMAGE DECODING AND COMPOSING APPARATUS

FIELD OF THE INVENTION

The present invention relates to a video image decoding and composing method which can realize interactive operation by a user and a video image decoding and composing apparatus, and also a video image composition information coding method.

BACKGOUND OF THE INVENTION

Up to now, video image compression method such as MPEG1, or MPEG2 has been utilized in the coding of natural video images. Further, there is an object coding method which divides a moving video image into objects in the video image and encodes the video images for each object and each background, as a new coding method.

Fig. 7 is a conceptional view showing composition of object video images. In the figure, numeral 701 represents a background video image, and numerals 702 to 705 designate video images of objects in the video images. Numerals 706 and 707 designate composed images. As the objects, there are rectangular shape object images such as background image 701 and arbitrary shape objects having arbitrary shapes other than rectangular shape such as objects 702 to 705. The arbitrary shape object is constituted by a texture video image showing a color signal and a shape video image showing the shape. The objects of the video images 704, 705 are supposed to be located in the foreground relative to the objects of the video images 702, 703.

First of all, the background video image 701 and the texture video

image 702 of the object which is the closest to the background are composed using the shape video image 703, thereby to output the composed video image 706. Next, the texture video image 704 is composed with the composed video image 706 using the shape video image 705, thereby to output the composed video image 707. By these operations, the composed video image is produced. Herein, as shape video images, there are a two-value video image which only indicates whether it is inside or outside of the object, and a video image which indicate the ratio of composition of the pixel values of the background to that of the object by multi-values, thereby enabling semi-transparent composition. In the object coding method, the rectangular shape object video image (701), the arbitrary shape object video image (702, 703), and the arbitrary shape object video image (704, 705) can be individually coded for each object. In the MPEG-4 video image coding method, it is possible to perform encoding of such object image having arbitrary shapes other than the rectangular shape.

On the other hand, even in the computer graphics data, the standardization of coding method has been advancing. As a standard coding method, there is a virtual reality modeling language. In this coding method, information of a top, a line, and a surface, and their materials (such as color, light reflection parameter) can be coded, and it is possible in the decoding apparatus to re-structure the scene of computer graphics by decoding the coded signal of the virtual reality modeling language.

Recently, a coding method comprising combination of the object

coding method and the computer graphics data coding method has also attracted attentions. When the computer graphics data coding method is extended to the composition of the object video image, changing of the composition positions and the composition with the computer graphics images, of the object images of the object coding method can be performed.

In MPEG-4, there is also realized a coding method comprising a combination of the above-described object coding method and the computer graphics data coding method. By extending the computer graphics data coding method to the composition of the object video image, it is possible to compose the object video image of the object coding method with the computer graphics video image. Thereby, it is possible to realize the computer graphics of higher presentation ability with relative to the prior art.

Fig.8 shows an example of a video image decoding and composing apparatus comprising a combination of the object coding method and the computer graphics data coding method. Hereinafter, a coded signal describing the composition information of the object video image as the above-described extended computer graphics data coded format is called as a composition information coded signal.

Numeral 801 designates a composition information coded signal, numeral 802 designates a composition information coded signal decoding means for analyzing the composition information coded signal 801 and outputting the result as composition information, numeral 803 designates a composition information memory storing the

composition information as the output of the composition information coded signal decoding means 802, numeral 804 designates a coded signal of the arbitrary shape object video image, numeral 805 designates an arbitrary shape object decoding means for decoding the coded signal 804, and numeral 806 designates a shape memory storing a shape video image signal which is decoded by the arbitrary shape object decoding means 805. Numeral 807 designates a texture memory for storing a texture video image signal which is decoded by the arbitrary shape object decoding means 805. Numeral 808 designates a coded signal of the rectangular shape object video image and numeral 809 designates a rectangular shape object decoding means for decoding the coded signal 808. Numeral 810 designates a video image memory for storing the video image signal which is decoded by the rectangular shape object decoding means 809. Numeral 811 designates a composing means for composing the shape signal stored in the shape memory 806, the texture signal stored in the texture memory 807, and the video image signal stored in the video image memory 810 in accordance with the composition information stored in the composition information memory 803. Numeral 812 designates a composition video image signal which is output from the composing means 811.

The operation of the video image decoding and composing apparatus constructed as described above will be described with reference to the drawings and the tables as follows:

An example of the composition information coded signal 801 is shown in Table 4. This is described in a format similar to that of

the virtual reality modeling language. For the detail of the format of the virtual reality modeling language, please see "VRML2.0 -3D cyber space structuring language-" by Kenjiro Miura, Asakura Shoten, 1996. In the format, a node and a field accompanying the node are included. In this example, Group, Shape, Appearance, MovieTexture, Rectangle are nodes. The "children" is a field of Group node, the "appearance" and "geometry" are fields of Shape node, the "texture" is a field of Appearance node, and url is a field of MovieTexture node. The Group node represents a group of nodes, and describes the collection of nodes at the "children" field. The "MovieTexture" node represents a moving video image which is to be texture mapped to the object (in this example, Rectangle node) which is represented by the "geometry" field of the Shape node, and the location of the coded video image signal corresponding to the moving video image is described in the url field.

The composition information coded signal may be compressed in the text format as shown in Table 4, or may be further compressed in the binary format as in MPEG-4.

The arbitrary shape object decoding means 805 inputs and decodes the arbitrary shape coded signal 804, and the decoded shape video image is stored in the shape memory 806 and the texture video image in the texture memory 807, respectively. The rectangular shape object decoding means 809 inputs and decodes the rectangular object coded signal 808, and the decoded video image is stored in the video image memory 810. The composing means 811 composes the texture video

image of the arbitrary shape object stored in the texture memory 807 and the rectangular shape video image stored in the video image memory 810 in accordance with the composition information stored in the composition information memory. When the arbitrary shape video image is composed, the shape video image in the shape memory 806 is used. The composing means 811 outputs the composed video image signal as the composed video image signal 812. It is possible for the composer of the coded data to perform the composition of the decoded video image object freely by using the composition information coded signal.

On the other hand, there is an attempt that a user of the display apparatus should perform an interactive operation for the display object by the computer graphics. In the above-described virtual reality modeling language, an interactive operation for the computer graphics object is realized.

However, up to now, while the interactive operation for the object of computer graphics has been devised, the interactive operation by the user for the arbitrary shape object video image in the object coding has not been conceived. For example in the case shown in Table 4, when the user intends to select the arbitrary shape object video image by an operation with such as a mouse, the shape information of the object video image is not considered. Therefore, the object video image may be erroneously selected although its position is out of the shape.

SUMMARY OF THE INVENTION

It is an object of the present invention to easily realize an

interactive operation by the user in a decoding and composing apparatus of an arbitrary shape object.

According to a first aspect of the present invention, a video image decoding and composing method for decoding a video image coded signal having shape information in shape units and composing the decoded signals together to decode an object video image, uses a shape video image obtained by decoding the video image coded signal in detecting an object video image relating to the object existing at a predetermined position on a video screen.

According to a second aspect of the present invention, a video image decoding and composing method for decoding a video image coded signal having shape information in shape units and composing the decoded signals together to decode an object video image, composes only respective shapes video image obtained by decoding the object video image as a video image signal, together.

According to a third aspect of the present invention, in the video image decoding and composing method as defined in the second aspect of the invention, said shape video images are composed after being geometrically transformed.

According to a fourth aspect of the present invention, in the video image decoding and composing method as defined in the second aspect of the invention, said shape video images are composed after being assigned colors in the shapes.

According to a fifth aspect of the present invention, in the video image decoding and composing method as defined in the second aspect

of the invention, said shape video images are composed after textures being mapped to the shape video images.

According to a sixth aspect of the present invention, a video image decoding and composing method for decoding a video image coded signal having shape information thereby to decode object video images in shape units, comprises a composition information coded signal indicating shape video images which are obtained when the video images are decoded being input, and a plurality of the video images being composed in accordance with said composition information coded signal.

According to a seventh aspect of the present invention, a video image decoding and composing apparatus comprises a composition information coded signal decoding means for decoding a composition information coded signal; a composition information memory for storing composition information as the output of said composition information coded signal decoding means; an arbitrary shape object decoding means for decoding a coded signal of an object video image having an arbitrary shape other than a rectangular shape; a shape memory for storing a shape signal which is decoded by the arbitrary shape object decoding means; a texture memory for storing a texture signal which is decoded by the arbitrary shape object decoding means; a rectangular shape object decoding means for decoding a coded signal of a rectangular shape object video image; a video image memory for storing a video image signal which is decoded by the rectangular shape object decoding means; a composing means for composing the shape

signal stored in the shape memory, the texture signal stored in the texture memory, and the video image signal stored in the video image memory in accordance with the composition information stored in the composition information memory; a position indicating means for indicating a position in a composed image which is output from the composing means; and a shape selecting means for selecting only the shape signal according to the instruction from the composing means.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram illustrating a video image decoding and composing apparatus according to a first embodiment of the present invention.

Fig. 2 is a conceptional view illustrating detection of the object video image in the first embodiment of the present invention.

Fig. 3 is a conceptional view illustrating the composition information memory in the first embodiment of the present invention.

Fig. 4 is a block diagram illustrating the composition information coded signal decoding means in the first embodiment of the present invention.

Fig. 5 is a conceptional view illustrating the generation of the shadow of the object video image in the first embodiment of the present invention.

Fig. 6 is a block diagram illustrating the composition information coding apparatus according to a third embodiment of the present invention.

Fig. 7 is a conceptional view illustrating composition of an

object video image.

Fig. 8 is a block diagram illustrating the prior art video image decoding and composing apparatus.

DETAILED DESCRIPTION OF THE PREFFERED EMBODIMENTS

An embodiment of the present invention will be described with reference to the drawings.

(Embodiment 1.)

First of all, the interactive operation which should be realized in this first embodiment will be described.

Fig.2 is a conceptional view illustrating the selection of the object. Reference numeral 201 designates a background video image (rectangular shape object video image), numeral 202 designates a texture video image of an arbitrary shape object video image, numeral 203 designates a shape video image corresponding to the texture video image 202, and numeral 204 designates a composed video image comprising the video images 201 to 203. It is supposed that the arbitrary shape object video images 202, 203 are positioned in the foreground. It is also supposed that a user pointed the position of 205. The decoding and composing apparatus searches the object which is positioned at 205 and it detects that the arbitrary shape objects 202 and 203 exists thereat. If the detected object is assumed to be an object which the user has selected, it is possible to perform the object selection by the user. By assigning the selection of the user to the arbitrary shape objects 202 and 203, it is possible to perform an interactive operation with the user. For example, the button

operation by the arbitrary shape object such that selection of a particular object by the user leads to calling out of a separate operation, or a drug operation of an arbitrary shape object by such as a mouse of a personal computer can be realized.

A video image decoding and composing apparatus according to the first embodiment of the present invention will be described with reference to Figures 1 to 5 and Tables 1 to 3.

Fig.1 is a block diagram illustrating an example of a basic structure of a video image decoding and composing apparatus according to the first embodiment of the present invention. In the figure, numeral 101 designates a composition information coded signal. Numeral 102 designates a composition information coded signal decoding means for decoding the composition information coded signal 101, numeral 103 designates a composition information memory storing composition information as the output of the composition information coded signal decoding means 102, numeral 104 designates a coded signal of an arbitrary shape object video image, numeral 105 designates an arbitrary shape object decoding means for decoding the coded signal 104, numeral 106 designates a shape memory storing a shape video image signal which is decoded by the arbitrary shape object decoding means 105. Numeral 107 designates a texture memory for storing a texture video image signal which is decoded by the arbitrary shape object decoding means 105. Numeral 108 designates a coded signal of a rectangular shape object video image and numeral 109 designates a rectangular shape object decoding means for decoding the coded signal

108. Numeral 110 designates a video image memory for storing a video image signal which is decoded by the rectangular shape object decoding means 109. Numeral 111 designates a composing means for composing the shape signal stored in the shape memory 106, the texture signal stored in the texture memory 107, and the video image signal stored in the video image memory 110 in accordance with the composition information stored in the composition information memory 103. Numeral 112 designates a composition video image signal which is output from the composing means 111. Numeral 113 designates a position indicating means for indicating a position in the output composed video images. Numeral 114 designates a shape selecting means for selecting only the shape signal in accordance with the indication by the composing means 111.

The operation of the video image decoding and composing apparatus constructed as described above will be described with reference to the drawings and the tables.

The composition information coded signal decoding means 102 inputs and decodes the composition information coded signal 101 and stored a decoded result in the composition information memory 103. The details of the composition information coded signal 101, the composition information coded signal decoding means 102, and the composition information memory 103 will be described later. The arbitrary shape object decoding means 105 inputs and decodes the arbitrary shape object video image coded signal 104 and stores a decoded shape video image in the shape memory 106 and a decoded texture

video image in the texture memory 107, respectively. The rectangular shape object decoding means 109 inputs and decodes the rectangular shape object video image coded signal 108 and stores a decoded video image in the video image memory 110. The composing means 111 composes the shape video image stored in the shape memory 106, the texture video image stored in the texture memory 107, and the video image stored in the video image memory 110 in accordance with the composed video image information stored in the composition information memory 103, and output the result as the composed video image signal 112.

Here, the composition information coded signal will be described.

Table 1 shows an example of the composition information coded signal. This is described in the format similar to that of the virtual reality modeling language. Hereinafter, this Table 1 will be described with reference to figures 3 and 4.

Fig. 3 is a conceptional view showing the composition information obtained from the composition information coded signal in Table 1 and Fig. 4 is a detailed block diagram illustrating the composition information coded signal decoding means 102. Reference numeral 401 designates a node analyzing means selecting means for selecting an appropriate node analyzing means upon receipt of the composition information coded signal 101. Numerals 402 to 409 designate node analyzing means corresponding to the respective nodes of MovieTexture node, Bitmap node, TouchSensor node, Shape node, Transform node, Appearance node, Material node, and Group node. Numerals 410 to 417 designate node tables for storing the composition information as the

analysis results of the respective node analysis means 402 to 409. The Group, Shape, TouchSensor nodes are the same as those in the Virtual Reality Modeling Language (for detail grammar of the virtual reality modeling language, please see "VRML2.0 -3D cyber space structuring language-" By Kenjiro Miura, Asakura Shoten, 1996). With reference to the above drawings, the composition information of Table 1 will be described. Here, the names of the nodes are assigned for convenience, and if the nodes have the same functions, difference names may be assigned.

The TouchSensor node in the Table 1 is a node for performing a different processing when the user selects a shape which is related to the TouchSensor node by clicking with such as a mouse, and the interactive operation with the user can be described in this node. When the shape which is related to the TouchSensor node is selected, the TouchSensor node is made effective (The processing when the TouchSensor node is effective, is of course described in the TouchSensor node, but it is omitted in the description here for simplification). In this first embodiment, the interactive operation using the TouchSensor node will be described.

By the Group node on the first line in Table 1, the node analyzing means selecting means 401 selects the Group node analyzing means 409. The Group node analyzing means 409 detects the children field. The children field includes the Shape node. This information can be stored by the Group node analyzing means 409 as in the Group node table 301. Similar selecting of the node analyzing means and

analyzing processing are performed for the each node in the children field of the Group node. After all the nodes are analyzed, the composition information as shown in fig. 3 is obtained. The arrows indicate the references to the different node tables and these references are held as a memory pointer or an identifier to the reference destination node tables.

The composition information coded signal according to the present invention is different from the prior art in that there exists the Bitmap node for indicating the shape of the object video image signal. In this example, the Bitmap node refers the shape signal included in the video image coded signal described in the url field of the MovieTexture node. This shape is related to the TouchSensor node, and the when the position in the shape is selected, the TouchSensor node is made effective.

The interactive operation for the arbitrary shape object video image when the TouchSensor node is utilized will be described. The user indicates a position in the video image by the position indicating means 113 such as a mouse. The object selection will be judged from the objects positioning in the foreground successively as follows. A shape in the shape memory of the object video image to be judged is selected by the shape selection means 114, a pixel value of the position in the video image indicated is searched, and if the pixel value shows that the pixel is in the object shape, then the object video image is supposed to be selected. When there exists a TouchSensor node which is related to the selected object video image,

the TouchSensor node is made effective. If the object video image positioning in the foreground is not selected, the judgment processing is carried out repeatedly sequentially from the object video image in the foreground up to the background video image.

As described above, in this first embodiment, by defining the node indicating the shape of the object video image and using that node, the interactive operation by the user in the video image decoding and composing apparatus which decodes and composes the arbitrary shape object can be realized.

In this first embodiment, the Bitmap node which can refer to the shape from the video image object is defined, but the definition of the TouchSensor node can be extended as shown in the following.

Table 2 is an example of a composition information coded signal including the extended TouchSensor node. The extended TouchSensor node is related to the shapes included in all the children nodes of the parent node for the TouchSensor node similarly to the usual TouchSensor node. In the prior art method, the TouchSensor node on line 3 of Table 2 is related to the shape of the Rectangle on line 10 and when the object shown by the Rectangle node is selected, the TouchSensor node is made effective. However, when the shape is included in the object video image which is referred to in the url field of the MovieTexture node, the TouchSensor node should not be made effective when it is outside the shape. In this way, it is possible to judge whether the object video image exists at the selected position by the similar method to that when the Bitmap node is

utilized.

In addition, in the first embodiment, the TouchSensor node is utilized for the interactive operation, but the other sensor nodes such as CylinderSensor node, SphereSensor node, and PlaneSensor node in the virtual reality modeling language can be used therefor. The functions of these sensor nodes may be extended for the coded signals of the arbitrary shape object similarly to the TouchSensor in the first embodiment and the nodes whose functions are extended may be defined as new nodes.

While in this embodiment the definition of the TouchSensor is extended, a node which has the equivalent function to that of the extended TouchSensor can be defined as a different (separate) new node.

(Embodiment 2.)

While in the first embodiment the function of the TouchSensor node for the shape video image is described, the shape video image itself can be utilized for the display. More particularly, it is possible to perform the geometrical transformation, addition of color, and attaching of texture to the shape video image to realize the composition display. Hereinafter, the production of shadow of the object image using the shape video image will be described in this second embodiment.

Fig. 5 is a conceptional view showing the production of shadow of the object vide image. Numeral 501 designates a background video image and numeral 502 designates a texture video image of the object.

Numeral 503 designates a shape video image of the object. Numeral 504 designates a shape video image which is produced by transforming the shape video image 503. Numerals 505 and 506 designate composed video images, respectively. The shape video image 503 is transformed into the shape video image 504 showing the shadow thereof, by an appropriate geometrical transformation. The shape video image 504 is composed with the background video image 501 to produce the composition video image 505. The composition is performed supposing the pixel value in the shape of the video image 504 (pixel value of the shadow) has a constant value in place of using the texture video image 502. The composition video image 505 and the arbitrary shape object video images 502, 503 are composed to output the composed video image 506.

Table 3 shows an example of composition information coded signal when producing a shadow. The Group, Transform, Shape, Appearance, Material nodes are the same as those in the virtual reality modeling language. In the url field of the Bitmap node, there is described the location of a coded signal of the object video image ("video1"). The material characteristics, i.e., color of the shape is given by the Material node on lines 9 to 11. The shape to which the color is attached is geometrically transformed by the Transform node, and the shadow of the arbitrary shape object video image is generated, and the video image which is obtained by composing the shadow and the arbitrary shape object video image is output. When a texture is to be attached to the shape, the texture may be assigned to the texture

field of the Appearance node.

Here, the video image signal in this embodiment may be a still video image or a moving video image.

In addition, as the arbitrary shape object video image coded signal in the second embodiment, the video image coded signal indicating the transparent pixel value as the transparent type GIF file and the transparent type animation GIF file may be employed. In this case, the object shape can be obtained depending on whether the pixel value of the video image is transparent or not, similarly in the processing in this embodiment.

Further, in this embodiment, the arbitrary shape object video image signal which is obtained by coding the texture video image and the shape video image is input and decoded, but the coded signal of only the shape video image may be input and decoded.

Further, in this embodiment, when only the shape video image of the arbitrary shape object coded signal is referred, only the shape video image may be decoded.

Further, while in this embodiment the format using the url field of the Bitmap node is employed to indicate the location of the coded signal including the shape signal of this embodiment, it is, for example, possible to employ a format in which the MovieTexture node is used and the location of the coded signal is described in the url field therein, as in the first embodiment. Also in this case, if the texture is replaced by the material characteristics (color) described in the Material node, the same effects as those in this embodiment can be

obtained.

In addition, while in the first and second embodiments there are provided one arbitrary shape object decoding means and one rectangular shape object decoding means, a plurality of the respective means may be provided. Or there may be provided only one selected from the arbitrary shape object decoding means and the rectangular shape object decoding means.

In addition, the composition information coded signal of the first and second embodiments may be coded by in the letter format as shown in Table 1 or in the binary format.

In addition, in the first and second embodiments the format in the virtual reality modeling language is referred as the format of the composition information coded signal, but if the function indicating the shape of the object video image such as the function corresponding to the Bitmap node or the extended TouchSensor node in this embodiment can be represented, the other formats can be employed.

In addition, while in the first and second embodiments the composition information coded signal, the arbitrary shape object coded signal, and the rectangular shape coded signal are separately input, the multiplexed data may be input and de-multiplexed to the respective data to be processed.

(Embodiment 3.)

This third embodiment shows a video image coding and composition information coding apparatus which generates the composition

information coded signal, the arbitrary shape object video image coded signal, and the rectangular shape object video image coded signal, which are input in the first embodiment.

Fig. 6 is a block diagram illustrating an example of the fundamental construction of the video image coding and composition information coding apparatus according to the third embodiment of the present invention. In the figure, reference numeral 601 designates a composition information, numeral 602 designates composition information coding means for coding the composition information 601, numeral 603 designates a composition information coded signal which is output from the composition information coding means 602. Numeral 604 designates a shape video image signal. Numeral 605 designates a texture video image signal. Numeral 606 designates an arbitrary shape object coding means for coding the shape video image signal 604 and the texture video image signal 605. Numeral 607 designates an arbitrary shape object video image coded signal output from the arbitrary shape object coding means 606. Numeral 608 designates a rectangular shape video image signal. Numeral 609 designates a rectangular shape object coding means for coding the rectangular shape object video image signal 608. Numeral 610 designates a rectangular shape object video image coded signal which is output from the rectangular shape object coding means 609.

The operation of the video image coding and composition information coding apparatus constructed as described above will be described with reference to the drawings.

The composition information coding means 602 inputs and encodes the composition information 601 and outputs the composition information coded signal 603. As the format of the composition information coded signal 603, the format employed in the first embodiment can be employed. The node indicating the shape of the object video image can be coded by the composition information coding means 602. The arbitrary shape object coding means 606 inputs and encodes the texture video image signal 604 and the shape video image signal 605 corresponding thereto, and outputs the arbitrary shape object video image coded signal 607. The rectangular shape object coding means 609 inputs and encodes the rectangular shape object video image signal 608 to output the rectangular shape object video image coded signal 610.

As described above, in this embodiment, the composition information including the node indicating the shape of the video image object can be coded.

Here, while in the third embodiment there are provided one arbitrary shape object coding means and one rectangular shape object coding means, a plurality of the respective means may be provided. Or there may be provided only one selected from the arbitrary shape object coding means and the rectangular shape object coding means.

In addition, as the arbitrary shape object video image coded signal in the third embodiment, the video image coded signal indication the transparent pixel value as the transparent type GIF file and the transparent type animation GIF file may be employed.

In the third embodiment, only the shape video image may be coded by the arbitrary shape object coding means to output the coded signal.

Further, while in this embodiment the composition information coded signal, the arbitrary shape object video image coded signal, and the rectangular shape coded signal are separately output, these may be output with multiplexed.

Here, the composition information coded signal, the arbitrary shape object video image coded signal, and the rectangular shape coded signal in the third embodiment can also be transmitted and recorded.

As described above, according to the present invention, a video image decoding and composing method in which the interactive operation by the user can be easily performed is realized.

Table 1

Group {	1
children [	2
TouchSensor{.....}	3
Shape {	4
appearance Appearance {	5
texture MovieTexture	6
url "video1"	7
}	8
}	9
geometry Bitmap { }	10
}	11
]	12
}	13

Table 2

Group {	1
children [	2
TouchSensor{.....}	3
Shape {	4
appearance Appearance {	5
texture MovieTexture	6
url "video1"	7
}	8
}	9
geometry Rectangle {}	10
}	11
]	12
}	13

Table 3

Group {	1
children [	2
Transform {	3
scale {.....}	4
translation {...}	5
children [	6
Shape {	7
appearance Appearance {	8
material Material {	9
diffusecolor 0,0,0	10
}	11
}	12
geometry Bitmap {	13
url "video1"	14
}	15
}	16
]	17
}	18
]	19
}	20

Table 4

Group {	1
children [	2
Shape {	4
appearance Appearance {	5
texture MovieTexture	6
url "video1"	7
}	8
}	9
geometry Rectangle {}	10
}	11
]	12
}	13



WHAT IS CLAIMED IS:

1. A video image decoding and composing method for decoding a video image coded signal having shape information in shape units and composing the decoded signals together to decode an object video image, comprising:

a shape video image obtained by decoding the video image coded signal being used in detecting an object video image relating to the object existing at a predetermined position on a video screen.

2. A video image decoding and composing method for decoding a video image coded signal having shape information in shape units and composing the decoded signals together to decode an object video image, comprising:

only each of shapes video images obtained by decoding the object video image being composed together as a video image signal.

3. The video image decoding and composing method as defined in claim 2, wherein said shape video images is composed after the same being geometrically transformed.

4. The video image decoding and composing method as defined in claim 2, wherein said shape video images is composed after the same being assigned colors in the shapes.

5. The video image decoding and composing method as defined in claim 2, wherein said shape video images is composed after textures being mapped to the shape video images.

6. A video image decoding and composing method for decoding a video image coded signal having shape information thereby to decode

* C L M S *

object video images in shape units, comprising:

a composition information coded signal indicating shape video images which are obtained when the video images are decoded being input, and;

a plurality of the video images being composed in accordance with said composition information coded signal.

7. A video image decoding and composing apparatus comprising:

a composition information coded signal decoding means for decoding a composition information coded signal;

a composing information memory for storing composition information as the output from said composition information coded signal decoding means;

an arbitrary shape object decoding means for decoding a coded signal of an object video image having an arbitrary shape other than a rectangular shape;

a shape memory for storing a shape signal which is decoded by the arbitrary shape object decoding means;

a texture memory for storing a texture signal which is decoded by the arbitrary shape object decoding means;

a rectangular shape object decoding means for decoding a coded signal of a rectangular shape object video image;

a video image memory for storing a video image signal which is decoded by the rectangular shape object decoding means;

a composing means for composing the shape signal stored in the shape memory, the texture signal stored in the texture memory,

* C L M S *

and the video image signal stored in the video image memory in accordance with the composition information stored in the composition information memory;

a position indicating means for indicating a position in a composed video image which is output from the composing means; and

a shape selecting means for selecting only the shape signal according to the instruction from the composing means.