

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
27 December 2002 (27.12.2002)

PCT

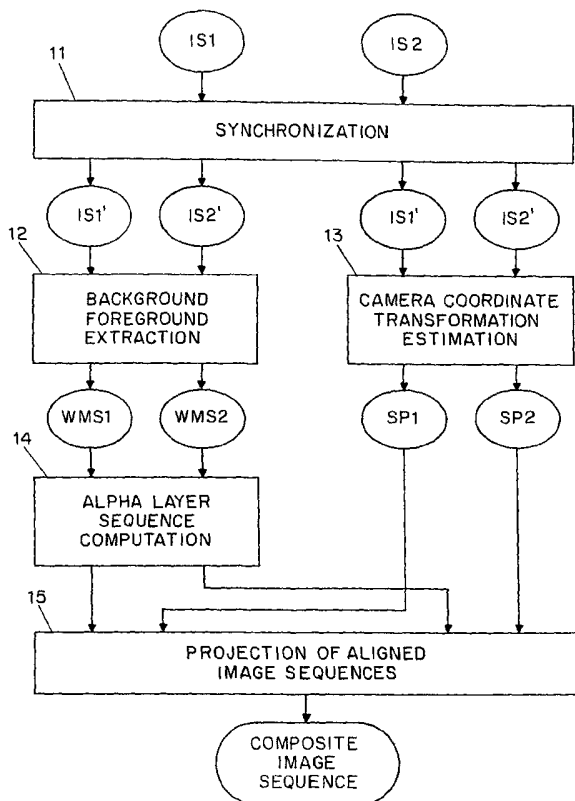
(10) International Publication Number  
**WO 02/104009 A1**

- (51) International Patent Classification<sup>7</sup>: **H04N 5/272**
- (21) International Application Number: PCT/US01/19741
- (22) International Filing Date: 19 June 2001 (19.06.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (71) Applicants: **ECOLE POLYTECHNIQUE FEDERALE DE LAUSANNE (EPFL)** [CH/CH]; c/o SRI, CH-1015 Lausanne (CH). **BUSINGER, Peter, A.** [US/US]; 344 Valleyscent Avenue, Scotch Plains, NJ 07076 (US).
- (72) Inventors: **PRANDONI, Paolo**; Avenue de Morges 16, CH-1004 Lausanne (CH). **VETTERLI, Martin**; Chemin de Baussan 11, CH-1091 Grandvaux (CH). **AYER, Serge**; Impasse Pra-Michaud 5, CH-1753 Matran (CH).
- (74) Agent: **BUSINGER, Peter, A.**; 344 Valleyscent Avenue, Scotch Plains, NJ 07076 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**  
— with international search report

[Continued on next page]

(54) Title: METHOD AND SYSTEM FOR COMBINING VIDEO WITH SPATIO-TEMPORAL ALIGNMENT



(57) Abstract: Given two video sequences (IS1, IS2), a composite video sequence can be generated which includes visual elements from each of the given sequences, suitably synchronized and represented in a chosen focal plane (figure 1). For example, given two video sequences with each showing a different contestant individually racing the same down-hill course, the composite representation can include elements from each of the given sequences to show the contestants as if racing simultaneously. Generating a composite representation can be facilitated by referring to a common background against which foreground objects are imaged. The composite representation can be formed by suitably blending together of foreground and background, and with re-scaling and/or re-framing for optimized presentation.

WO 02/104009 A1



---

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

## METHOD AND SYSTEM FOR COMBINING VIDEO WITH SPATIO-TEMPORAL ALIGNMENT

### Technical Field

The present invention relates to visual displays and, more specifically, to time-  
5 dependent visual displays.

### Background of the Invention

In video displays, e.g. in sports-related television programs, special visual effects  
can be used to enhance a viewer's appreciation of the action. For example, in the case of  
a team sport such as football, instant replay affords the viewer a second chance at  
10 "catching" critical moments of the game. Such moments can be replayed in slow motion,  
and superposed features such as hand-drawn circles, arrows and letters can be included  
for emphasis and annotation. These techniques can be used also with other types of  
sports such as racing competitions, for example.

With team sports, techniques of instant replay and the like are most appropriate,  
15 as scenes typically are busy and crowded. Similarly, e.g. in the 100-meter dash  
competition, the scene includes the contestants side-by-side, and slow-motion  
visualization at the finish line brings out the essence of the race. On the other hand,  
where starting times are staggered e.g. as necessitated for the sake of practicality and  
safety in the case of certain sports events such as ski jumping and downhill racing, the  
20 actual scene typically includes a single contestant. The invention described below  
includes aspects which are of interest to individual as well as team sports.

### Summary of the Invention

For enhanced visualization, by the sports fan as well as by the contestant and his  
coach, displays are desired in which the element of competition between contestants is  
25 manifested. This applies especially where contestants perform sole as in downhill skiing,  
for example, and can be applied also to group races in which qualification schemes are  
used to decide who will advance from quarter-final to half-final to final.

We have recognized that, given two or more video sequences, a composite video

sequence can be generated which includes visual elements from each of the given sequences, suitably synchronized and represented in a chosen focal plane. For example, given two video sequences with each showing a different contestant individually racing the same down-hill course, the composite sequence can include elements from each of the  
5 given sequences to show the contestants as if racing simultaneously. Further applications are in track and field, diving, horseback riding and golf, for close comparison between contestants. A composite video sequence can be made also by similarly combining one or more video sequences with one or more different sequences such as audio sequences, for example.

10 We have recognized further that generating a composite representation from given images can be facilitated by referring to a common background against which foreground objects are imaged. The composite representation can be formed by suitably blending together of foreground and background, and with re-scaling and/or re-framing for optimized presentation.

#### 15 Brief Description of the Drawing

Fig. 1 is a block diagram for synchronization and alignment processing.

Figs. 2A and 2B are schematics of different downhill skiers passing before a video camera.

20 Figs. 3A and 3B are schematics of images recorded by the video camera, corresponding to Figs. 2A and 2B.

Fig. 4 is a schematic of Figs. 2A and 2B combined.

Fig. 5 is a schematic of the desired video image, with the scenes of Fig. 3A and 3B projected in a chosen focal plane.

25 Fig. 6 is a frame from a composite video sequence which was made with a prototype implementation of the invention.

Fig. 7 is a block diagram of processing for building a database of sequences registered on a manifold.

Fig. 8 is a block diagram for re-framing in generating a target sequence.

Fig. 9 is a schematic illustrating spatial indexing on a cylinder.

### Detailed Description

A video sequence may be defined as a sequence of image fields, each containing the visual information pertaining to the portion of the physical world seen by a camera at contiguous, discrete time instants. For each field there is spatial information which is  
5 related to the imaged physical features, and to the instantaneous conditions of the camera such as displacement, position, aperture, pan angle and tilt angle, for example. Such information can be understood as a composite spatial index, representing a virtual camera frame and corresponding to a unique finite region of a multidimensional indexing manifold. Conversely, from a selected finite region of the manifold a spatial index can be  
10 derived which need not be unique.

Each field of the video sequence can be indexed further based on temporal information including a time index associated with each field, which can be used for temporal synchronization of different sequences. Combined spatial and temporal indexing can be interpreted as being on a multidimensional manifold which is used as a  
15 global coordinate system for all possible frames which can be captured by the camera within the range of its physical limitations.

Conceptually, the invention can be appreciated in analogy with 2-dimensional (2D) "morphing", i.e. the smooth transformation, deformation or mapping of one image, I1, into another, I2, in computerized graphics. Such morphing leads to a video sequence  
20 which shows the transformation of I1 into I2, e.g., of an image of an apple into an image of an orange, or of one human face into another. The video sequence is 3-dimensional, having two spatial and a temporal dimension. Parts of the sequence may be of special interest, such as intermediate images, e.g. the average of two faces, or composites, e.g. a face with the eyes from I1 and the smile from I2. Thus, morphing between images can be  
25 appreciated as a form of merging of features from the images.

The invention is concerned with a more complicated task, namely the merging of two video images, especially in video sequences. The morphing or mapping from one sequence to another leads to 4-dimensional data which cannot be displayed easily. However, any intermediate combination, or any composite sequence leads to a new video  
30 sequence.

Of particular interest is the generation of a new video sequence combining elements from two or more given sequences, with suitable spatio-temporal alignment or synchronization, and projection into a chosen focal plane. For example, in the case of a sports racing competition such as downhill skiing, video sequences obtained from two  
5 contestants having traversed a course separately can be time-synchronized by selecting the frames corresponding to the start of the race. Alternatively, the sequences may be synchronized for coincident passage of the contestants at a critical point such as a slalom gate, for example.

The chosen focal plane may be the same as the focal plane of the one or the other  
10 of the given sequences, or it may be suitably constructed yet different from both.

Of interest also is synchronization based on a distinctive event, e.g., in track and field, a high-jump contestant lifting off from the ground or touching down again. In this respect it is of further interest to synchronize two sequences so that both lift-off and touch-down coincide, requiring time scaling. The resulting composite sequence affords a  
15 comparison of trajectories.

With the video sequences synchronized, they can be further aligned spatially, e.g. to generate a composite sequence giving the impression of the contestants traversing the course simultaneously. In a simple approach, spatial alignment can be performed on a frame-by-frame basis. Alternatively, by taking a plurality of frames from a camera into  
20 consideration, the view in an output image can be extended to include background elements from several sequential images.

Forming a composite image involves representing component scenes in a chosen focal plane, typically requiring a considerable amount of computerized processing, e.g. as illustrated by Fig. 1 for the special case of two video input sequences.

25 Fig. 1 shows two image sequences IS1 and IS2 being fed to a module 11 for synchronization into synchronized sequences IS1' and IS2'. For example, the sequences IS1 and IS2 may have been obtained for two contestants in a down-hill racing competition, and they may be synchronized by the module 11 so that the first frame of each sequence corresponds to its contestant leaving the starting gate.

30 The synchronized sequences are fed to a module 12 for background-foreground

extraction, as well as to a module 13 for camera coordinate transformation estimation. For each of the image sequences, the module 12 yields a weight-mask sequence (WMS), with each weight mask being an array having an entry for each pixel position and differentiating between the scene of interest and the background/foreground. The  
5 generation of the weight mask sequence involves computerized searching of images for elements which, from frame to frame, move relative to the background. The module 13 yields sequence parameters SP1 and SP2 including camera angles of azimuth and elevation, and camera focal length and aperture among others. These parameters can be determined from each video sequence by computerized processing including interpolation  
10 and matching of images. Alternatively, a suitably equipped camera can furnish the sequence parameters directly, thus obviating the need for their estimation by computerized processing. For example, a camera can include sensors for tilt, roll and pan angles and record their instantaneous readings along with each image taken.

The weight-mask sequences WMS1 and WMS2 are fed to a module 13 for  
15 “alpha-layer” sequence computation. The alpha layer is an array which specifies how much weight each pixel in each of the images should receive in the composite image.

The sequence parameters SP1 and SP2 as well as the alpha layer are fed to a module 15 for projecting the aligned image sequences in a chosen focal plane, resulting in the desired composite image sequence. This is exemplified further by Figs. 2A, 2B, 3A,  
20 3B, 4 and 5.

Fig. 2A shows a skier A about to pass a position marker 21, with the scene being recorded from a camera position 22 with a viewing angle  $\Phi(A)$ . The position reached by A may be after an elapse of  $t(A)$  seconds from A's leaving the starting gate of a race event.

25 Fig. 2B shows another skier, B, in a similar position relative to the marker 21, and with the scene being recorded from a different camera position 23 and with a different, more narrow viewing angle  $\Phi(B)$ . For comparison with skier A, the position of skier B corresponds to an elapse of  $t(A)$  seconds from B leaving the starting gate. As illustrated, within  $t(A)$  seconds skier B has traveled farther along the race course as compared with  
30 skier A.

Figs. 3A and 3B show the resulting respective images.

Fig. 4 shows a combination with Figs. 2A and 2B superposed at a common camera location.

Fig. 5 shows the resulting desired image projected in a chosen focal plane, affording immediate visualization of skiers A and B as having raced jointly for  $t(A)$  seconds from a common start.

Fig. 6 shows a frame from a composite image sequence generated by a prototype implementation of the technique, with the frame corresponding to a point of intermediate timing. The value of 57.84 is the time, in seconds, that it took the slower skier to reach the point of intermediate timing, and the value of +0.04 (seconds) indicates by how much he is trailing the faster skier.

The prototype implementation of the technique was written in the "C" programming language, for execution on a SUN Workstation or a PC, for example. Dedicated firmware or hardware can be used for enhanced processing efficiency, and especially for signal processing involving matching and interpolation.

Individual aspects and variations of the technique are described below in further detail.

#### A. Background/Foreground Extraction

In each sequence, background and foreground can be extracted using a suitable motion estimation method. This method should be "robust", for background/foreground extraction where image sequences are acquired by a moving camera and where the acquired scene contains moving agents or objects. Required also is temporal consistency, for the extraction of background/foreground to be stable over time. Where both the camera and the agents are moving predictably, e.g. at constant speed or acceleration, temporal filtering can be used for enhanced temporal consistency.

Based on determinations of the speed with which the background moves due to camera motion, and the speed of the skier with respect to the camera, background/foreground extraction generates a weight layer which differentiates between those pixels which follow the camera and those which do not. The weight layer will then



be used to generate an alpha layer for the final composite sequence.

#### B. Spatio-temporal Alignment of Sequences

Temporal alignment involves the selection of corresponding frames in the sequences, according to a chosen criterion. Typically, in sports racing competitions, this is the time code of each sequence delivered by the timing system, e.g. to select the frames  
5 corresponding to the start of the race. Other possible time criteria are the time corresponding to a designated spatial location such as a gate or jump entry, for example.

Spatial alignment is effected by choosing a reference coordinate system for each frame and by determining a camera coordinate transformation between the reference  
10 system and the corresponding frame of each sequence. Such a determination may involve position estimating based on image contents and/or measuring, e.g. using the global positioning system (GPS). A determination may be unnecessary when camera data such as camera position, viewing direction and focal length are recorded along with the video sequence. Typically, the reference coordinate system is chosen as one of the given  
15 sequences, namely the one to be used for the composite sequence. As described below, spatial alignment may be on a single-frame or multiple-frame basis.

##### B.1 Spatial Alignment on a Single-frame Basis

At each step of this technique, alignment uses one frame from each of the sequences. As each of the sequences includes moving agents/objects, the method for  
20 estimating the camera coordinate transformation needs to be robust. To this end, the masks generated in background/foreground extraction can be used. Also, as motivated for background/foreground extraction, temporal filtering can be used for enhancing the temporal consistency of the estimation process.

##### B.2 Spatial Alignment on a Multiple-frame Basis

25 In this technique, spatial alignment is applied to reconstructed images of the scene visualized in each sequence. Each video sequence is first analyzed over multiple frames for reconstruction of the scene, using a technique similar to the one for

background/foreground extraction, for example. Once each scene has been separately reconstructed, e.g. to take in as much background as possible, the scenes can be spatially aligned as described above.

This technique allows free choice of the field of view of every frame in the scene, in contrast to the single-frame technique where the field of view has to be chosen as the one of the reference frame. Thus, in the multiple-frame technique, in case that all contestants are not visible in all the frames, the field and/or angle of view of the composite image can be chosen such that all competitors are visible.

#### C. Superimposing of Video Sequences

After extraction of the background/foreground in each sequence and estimation of the camera coordinate transformation between each sequence and a reference system, the sequences can be projected into a chosen focal plane for simultaneous visualization on a single display. Alpha layers for each frame of each sequence are generated from the multiple background/foreground weight masks. Thus, the composite sequence is formed by transforming each sequence into the chosen focal plane and superimposing the different transformed images with the corresponding alpha weight.

#### D. Applications

Video displays are of interest for TV broadcasting as well as for on-demand services, for example. The latter may allow for user interaction, e.g. in choosing camera angle, zooming, choice of viewpoint, and choice of contestants whose performance a user may wish to compare. Such a service may be provided by an Internet-based sports site and may include enhancements such as graphing of virtual trajectories, marking of spatial locations for performance comparison among contestants, and stroboscopic of fast events which involves displaying an event as “frozen” in space, by a series of overlapping snapshots taken at short intervals of time.

Further to skiing competitions as exemplified, the techniques of the invention can be applied to other speed/distance sports such as car racing competitions and track and field, for example.

Further to visualizing, one application of a composite video sequence made in accordance with the invention is apparent from Fig. 6, namely for determining differential time between two runners at any desired location of a race. This involves simple counting of the number of frames in the sequence between the two runners passing the  
5 location, and multiplying by the time interval between frames.

A composite sequence can be broadcast over existing facilities such as network, cable and satellite TV, and as video on the Internet, for example. Such sequences can be offered as on-demand services, e.g. on a channel separate from a strictly real-time main channel. Or, instead of by broadcasting over a separate channel, a composite video  
10 sequence can be included as a portion of a regular channel, displayed as a corner portion, for example.

In addition to their use in broadcasting, generated composite video sequences can be used in sports training and coaching. And, aside from sports applications, there are potential industrial applications such as car crash analysis, for example.

15 It is understood that composite sequences may be higher-dimensional, such as composite stereo video sequences.

In yet another application, one of the given sequences is an audio sequence to be synchronized with a video sequence. Specifically, given a video sequence of an actor or singer, A, speaking a sentence or singing a song, and an audio sequence of another actor,  
20 B, doing the same, the technique can be used to generate a voice-over or "lip-synch" sequence of actor A speaking or singing with the voice of B. In this case, which requires more than mere scaling of time, dynamic programming techniques can be used for synchronization.

The spatio-temporal realignment method can be applied in the biomedical field as  
25 well. For example, after orthopedic surgery, it is important to monitor the progress of a patient's recovery. This can be done by comparing specified movements of the patient over a period of time. In accordance with an aspect of the invention, such a comparison can be made very accurately, by synchronizing start and end of the movement, and aligning the limbs to be monitored in two or more video sequences.

30 Another application is in car crash analysis. The technique can be used for

precisely comparing the deformation of different cars crashed in similar situations, to ascertain the extent of the difference. Further in car crash analysis, it is important to compare effects on crash dummies. Again, in two crashes with the same type of car, one can precisely compare how the dummies are affected depending on configuration, e.g. of  
5 safety belts.

E. Further Synchronization and Alignment Considerations

Temporal indexing of video sequences can be performed in different ways, depending on their content and the blending goal. In sport races such as skiing, video sequences can be temporally synchronized according to their timing information, usually  
10 a chronometric counter which is reset at the start of each run. After synchronization, spatial alignment of the sequences results in a single video sequence showing at each instant in time the relative position of contestants, thus highlighting trajectory and speed differences between contestants. Alternatively, video sequences can be aligned with respect to spatially derived information, e.g. the instant of a competitor's crossing a pre-  
15 selected line in the environment as in high-jump and other track-and-field events. Some sequences may consist only of static background information, needing no temporal synchronization but only spatial indexing for use in blending. For example, background information may be provided by a camera scan of the empty race field, taken prior to the sport event.

20 Spatial indexing of video sequences can be effected by hardware or software. Hardware can include camera sensors which measure its instantaneous physical status and provide corresponding data along with the visual information in recorded fields. Software can provide for robust estimation techniques for inferring the relative displacement of the camera, from a sequence of recorded visual fields.

25 Fig. 7 illustrates how a database of sequences registered on a manifold can be constructed. A pre-processor 71 assembles video information 72 and camera parameters 73 if available. The assembled data and, unless a flag 74 is set to identify the video information 72 as static background information, synchronization information 75 are furnished to a manifold projection module 76 which produces a database update 77 for

the database 78.

Blending of temporally and spatially indexed sequences can be effected as follows: First the two or more original sequences selected for blending are analyzed in terms of their spatial indexing. From the indexing information, an extended background  
5 is reconstructed using the global indexing on the manifold of the background sequences. The original sequences are now synchronized, and a new target sequence is formed by spatially aligning the original, synchronized sequences with their common background on a field-by-field basis. Such spatial alignment can be effected by a robust camera motion estimation technique, e.g. including explicit indexing of each field on the global  
10 manifold. The final target sequence is obtained by blending the visual information of the original, spatio-temporally aligned sequences and of the background information on a field-by-field basis over a suitably defined viewing area. The viewing area can be determined automatically and so as to ensure that all objects of interest in the original sequences appear in the same field of view in the blended sequence. Special viewing  
15 needs can be accommodated by operator-controlled re-framing, for example.

Blending can be effected by reference to an established alpha layer, i.e. a relative weight array prescribing the relative contributions of each original field to the blended field. The alpha layer is determined based on information about the original sequences, e.g. the location of active foreground areas obtained by a robust foreground/background  
20 extraction method.

Series of target frames can be user-definable for effecting various video processing manipulations, e.g. slow motion and re-framing. For example, two ski race sequences which have little or no overlap with each other but share a common background can be integrated into a common field of view, based on background  
25 information.

For sports broadcasting for example, enhancements can be included such as virtual trajectories or reference lines embedded in the background information, so that such trajectories or lines automatically will be properly positioned when the background is combined with an event sequence. Facilitated further is the generation of stroboscopic  
30 still images including several images of an athlete in the course of his trajectory, e.g. in a

broad-jump event, by using background information and blending of camera fields selected according to their time index from the beginning of the jump.

### E.1 Spatial Indexing of a Video Sequence

Stills as captured by a video camera will be called video fields. The image captured in each field relates to the world around the camera via a set of parameters, including the geographical coordinates of the camera, three direction angles formed by the camera with respect to a chosen Cartesian reference system and usually called pan, tilt and roll angles, the camera aperture or zoom, and several physical parameters related to camera components, e.g. the lens, photosensitive elements and shutter speed. Some of these camera parameters are fixed, while others vary under control by the cameraman in the course of a shoot. When suitably equipped with sensors, a camera can furnish such parameters directly; otherwise they can be estimated computationally on the basis of motion characteristics of a recorded video sequence, using one of a number of known robust motion estimation techniques and mapping to the global reference system.

The camera-movement parameter values are delimited by mechanical camera limitations. The parameter ranges define a multidimensional manifold which can be used to spatially index a video sequence produced by the camera.

The use of spatial indexing for combining sequences can be visualized as being on a suitable projection surface or manifold, e.g. a cylindrical or spherical surface centered at the camera location. A region of the surface then represents a view from the camera position. Fig. 9 illustrates indexing on a cylinder, of two video sequences of frame regions 91, . . . , 91' and 92, . . . , 92'. Shown further is a region 93 which corresponds to a desired view for combining the video sequences. When suitably synchronized based on timing information, all those frame regions or portions thereof which overlap with the region 93 can contribute to a desired combined sequence. Conveniently, indexing on a cylinder involves recording azimuth and elevation. For indexing on a sphere, azimuth and declination can be used.

### E.2 Temporal Indexing of a Video Sequence

With each field of a video sequence, a temporal index can be associated for temporal alignment or synchronization between different sequences representing different events in the same environment. The index can be based on choice of a suitable starting instant. A sequence which only represents background will not require temporal indexing  
5 for blending with an action sequence.

### E.3 Extended Background Reconstruction

Background sequences can be spatially indexed on a suitably dimensioned manifold which can be taken as the reference system for the camera. Any other sequence, of an event against the background, can be projected on the same manifold to obtain a  
10 sequence of manifold coordinates which can be brought in correspondence with a series of fields in the indexed background sequences. The visual information in this series of background fields can be stitched together using a robust image stitching and mosaicing technique, defining an extended background image for the sequence. The width and extent of the background image can be modified readily. Image processing techniques  
15 can be applied to the background information prior to forming a target sequence, including the drawing of virtual trajectories and targets, color coding of image areas, and image enhancement among others.

### E.4 Realignment With Background Blending

Starting from an arbitrary number of original video sequences and their composite  
20 background image, a new target sequence is formed, composed of a number of contiguous target fields. This is effected by selecting, for each target field, a number of fields in the original sequences according to a chosen criterion. The visual information in the original fields is then suitably warped and blended with the visual information of their common reconstructed background to form each target field.

25 Spatial alignment can be effected by aligning the selected original fields with the reconstructed background. This operation relies on robust camera motion estimation software and/or hardware and may employ the same spatial indexing techniques as described above. A reference system is selected for the common background, serving as

multidimensional manifold as described above, and the original sequence frames are mapped onto this reference system by means of suitable warping techniques.

Synchronization is achieved by selecting those fields of the original sequences whose time indices, match a desired target time index. Suitable criteria include:

- 5 (i) Spatio-temporal realignment of two or more sequences with extended background reconstruction, with one of the original sequences chosen as a reference sequence. For each field in the reference sequence the target indices are computed by selecting the fields in the other original sequences so that their time indices match. The selected fields are then spatially aligned with their reconstructed background visual  
10 information.
- (ii) Temporal realignment of one or more sequences with extended background reconstruction for slow motion, wherein, for each field in the target sequence, an arbitrary target time index is computed, and all those fields in the original sequences are selected whose time index matches the target index. The selected fields are then spatially aligned  
15 with their reconstructed background visual information.

#### E.5 Blending

Once a set of target fields has been determined together with their common background information, blending can be effected by processing as follows:

- (i) *Object Motion Estimation*. For each original sequence a  
20 background/foreground extraction is performed, using a robust background-foreground estimation method. Robustness is called for here and throughout in the interest of processing image sequences acquired with a moving camera and containing moving persons and/or objects. Similarly called for is temporal consistency, i.e. foreground-background extraction is stable over time. As both the camera and people are moving  
25 according to physical properties, e.g. constant speed or acceleration, temporal filtering can be used for improving temporal consistency. Background-foreground extraction is aimed at generating a weight layer for distinguishing the portions of the original fields which follow the camera motion from those which do not. The weight layer will then be used in generating an alpha layer for the final composite sequence.



(ii) *Selection of the Viewing Area.* For each target field, a viewing area is defined on the extended reconstructed background according to a chosen viewing criterion.

Suitable criteria include:

(a) Multiple blending for trajectory comparison and stroboscoping, wherein the viewing area is defined for each target frame, suitably sized to encompass the visual information of the selected original frames after alignment with the background;

(b) Re-framing and virtual camera motions, wherein the viewing area is a user-defined variable which allows to define new camera trajectories over the common background;

(c) Alpha layer blending, wherein an alpha layer, i.e. an array of weight coefficients, is defined for the target frame, taking into account the results of the object motion estimate and the difference between original sequences and common background. The weights are used to combine the selected frames and the underlying background information via a weighted sum.

Fig. 8 illustrates processing for sequence re-framing with reference to a database 78 as generated according to Fig. 7, for example, and with reference further to the background flag 74, synchronization information 75, user-defined virtual camera parameters 81 and output parameters 82 which indicate the target type, e.g. as slow motion, stroboscopic or superposed. Frame retrieval modules 83 and 84 furnish frames to respective blending modules 85 and 86 which in turn forward the respective blended background and foreground sequences to a final blending module 87 for blending into the target sequence output.

Re-framing can be used to advantage further, to give an impression of zoom-in or zoom-out. This is of particular interest in case of motion along the line of sight, as in ski-jump events, for example.

Separately or in combination with re-framing, time re-scaling can stretch or compress time in an output video as compared with the original video, linearly or in any desired monotonic fashion. For example, for a more immediate comparison at critical points, of two participants in a triple-jump sports event for example, their videos can be synchronized so that their consecutive touch-downs appear as simultaneous.

CLAIMS

1. A method for generating a composite representation from a plurality of given images, comprising:
  - (a) synchronizing the given images;
  - 5 (b) relating the synchronized images to a common background; and
  - (c) forming the composite representation from the related images as projected into a chosen focal plane.
2. The method according to claim 1, wherein the composite representation comprises a video sequence.
- 10 3. The method according to claim 1, wherein the composite representation comprises a still image.
4. The method according to claim 1, wherein synchronizing is with respect to a timed event in the given images.
5. The method according to claim 1, wherein synchronizing is with respect to a  
15 common spatial event in the given images.
6. The method according to claim 1, wherein synchronizing is with respect to two events in each of the given images, with time scaling for equalizing time between the events.
7. The method according to claim 1, wherein the given images have camera  
20 parameters including camera location, orientation and focal length, wherein the chosen focal plane corresponds to the focal plane of the one of the given images, and wherein the composite representation is as viewed from the camera location and orientation of the one of the given images.

8. The method according to claim 1, wherein forming the composite representation is on a frame-by-frame basis in a video sequence.

9. The method according to claim 1, wherein forming the composite representation is based on a plurality of given images, for an expanded and/or re-framed  
5 view in the composite representation.

10. The method according to claim 1, wherein the given images are from a sports event.

11. The method according to claim 10, wherein the sports event comprises a ski race.

10 12. The method according to claim 10, wherein the sports event comprises a jump.

13. The method according to claim 10, wherein the sports event comprises a car race.

14. The method according to claim 1, wherein the common background is  
15 represented by a sequence of background images.

15. The method according to claim 14, further comprising establishing the background sequence prior to action sequences.

16. The method according to claim 15, wherein establishing the background sequence comprises video scanning the common background.

20 17. The method according to claim 14, further comprising image processing of background images.

18. The method according to claim 17, wherein image processing comprises including at least one virtual trajectory/target.
19. The method according to claim 17, wherein image processing comprises color coding.
- 5           20. The method according to claim 17, wherein image processing comprises image enhancement.
21. The method according to claim 1, wherein synchronizing comprises selecting fields in video sequences.
22. The method according to claim 21, wherein selecting the fields is for  
10 rendering critical points along trajectories as coincident.
23. The method according to claim 1, wherein forming the composite representation comprises blending of fields of the given images.
24. The method according to claim 23, wherein blending comprises background-foreground separation.
- 15           25. The method according to claim 23, wherein blending comprises referring to an alpha layer.
26. The method according to claim 1, wherein forming the composite representation comprises re-framing.
27. The method according to claim 26, wherein re-framing comprises a change in  
20 viewing direction.

28. The method according to claim 26, wherein re-framing comprises a change of scale.

29. The method according to claim 26, wherein re-framing is for trajectory comparison.

5           30. The method according to claim 26, wherein re-framing is for stroboscoping.

31. A system for generating a composite representation from a plurality of given images, comprising:

- (a) means for synchronizing the given images;
- (b) means for relating the synchronized images to a common background; and
- 10           (c) means for forming the composite representation from the related images as projected into a chosen focal plane.

32. A method for generating a composite representation from a plurality of given images, comprising:

- (a) synchronizing the given images;
- 15           (b) indexing the synchronized images on a manifold; and
- (c) forming the composite representation from images which share a common location on the manifold.

33. The method according to claim 32, wherein the manifold is cylindrical.

34. The method according to claim 32, wherein the manifold is spherical.

20           35. A system for generating a composite representation from a plurality of given images, comprising:

- (a) means for synchronizing the given images;
- (b) means for indexing the synchronized images on a manifold; and

(c) means for forming the composite representation from images which share a common location on the manifold.

36. A system for generating a composite representation from a plurality of given images, comprising:

- 5 (a) means for synchronizing the given images;  
(b) means for relating the given images to a common background; and  
(c) a processor which is instructed for forming the composite representation from the synchronized images as projected into a chosen focal plane.

37. A method for generating a composite image, comprising:

- 10 (a) synchronizing a plurality of given images;  
(b) relating the synchronized images to a common background; and  
(c) forming the composite image from the related images as projected into a chosen focal plane.

38. A broadcast service comprising:

- 15 (a) synchronizing a plurality of images;  
(b) relating the synchronized images to a common background;  
(c) forming a composite representation from the related images as projected into a chosen focal plane; and  
(d) broadcasting the composite representation.

20 39. An interactive service comprising:

- (a) accepting user input for synchronizing a plurality of images and synchronizing the images in accordance with the user input;  
(b) relating the synchronized images to a common background; and  
(c) forming a composite representation from the related images as projected into a  
25 chosen focal plane.

40. The service in accordance with claim 39, offered via a TV channel.

41. The service in accordance with claim 39, offered via Internet.

42. The service in accordance with claim 39, wherein user input comprises input for re-framing.

5           43. The service in accordance with claim 39, wherein user input comprises input for slow motion.

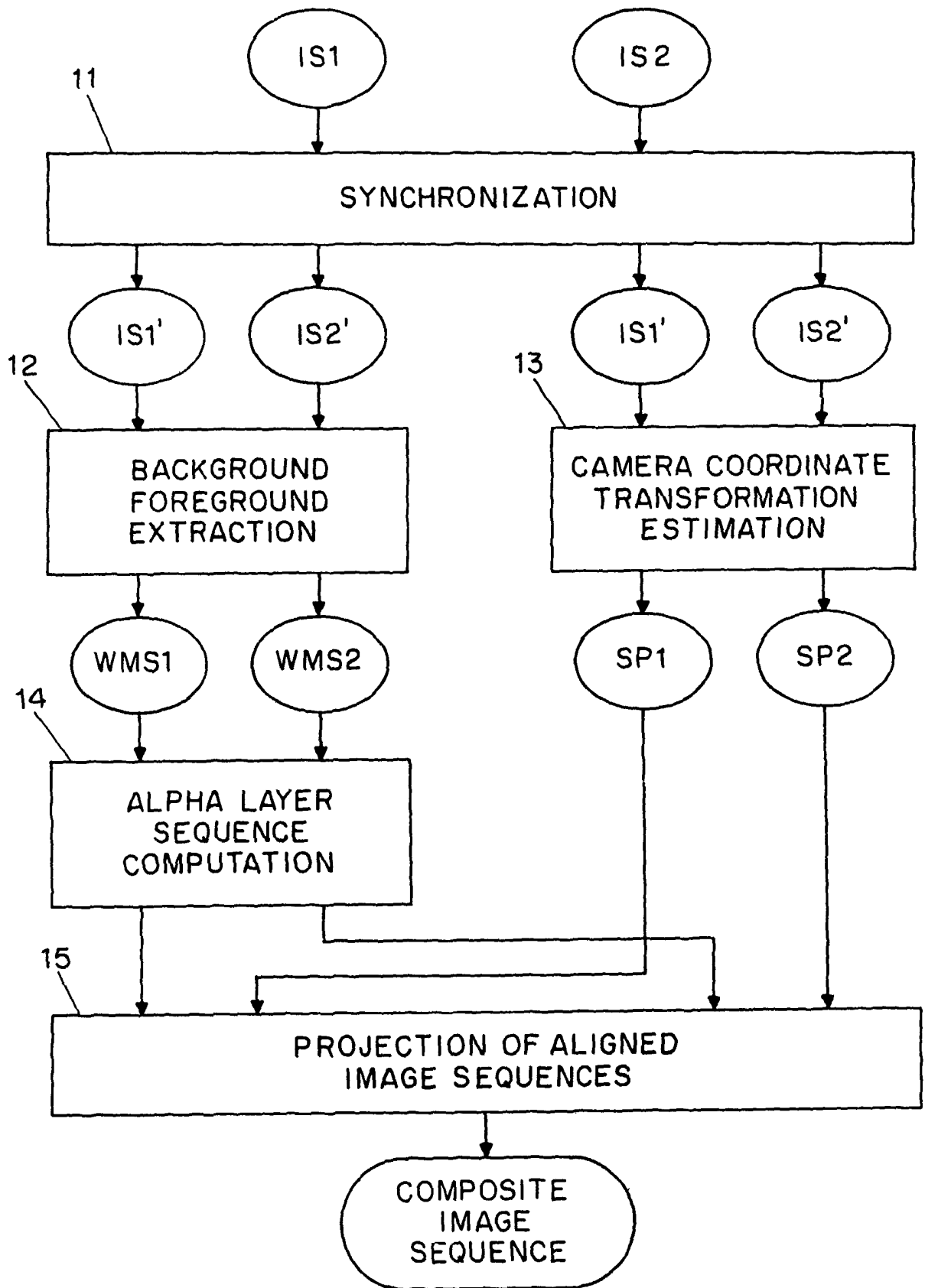


FIG. 1



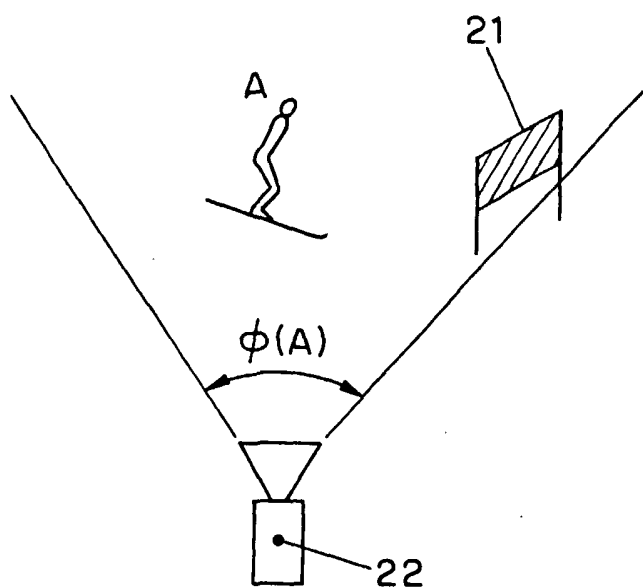


FIG. 2A

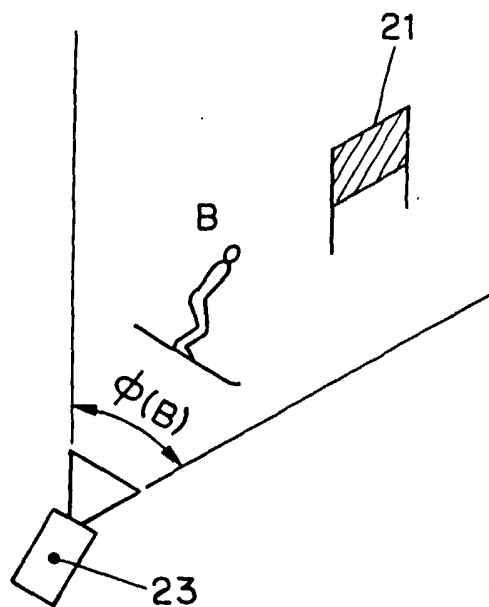


FIG. 2B

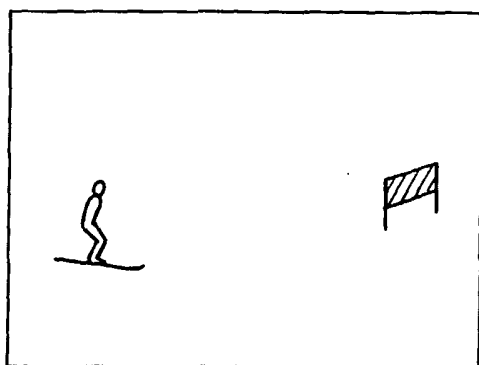


FIG. 3A

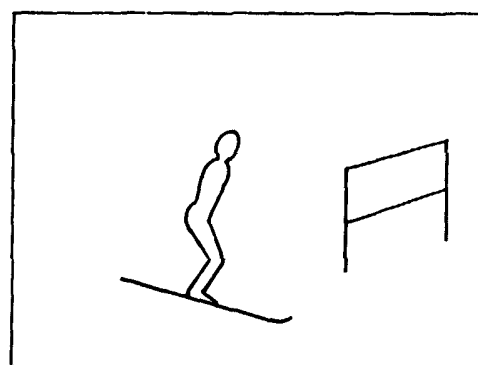


FIG. 3B

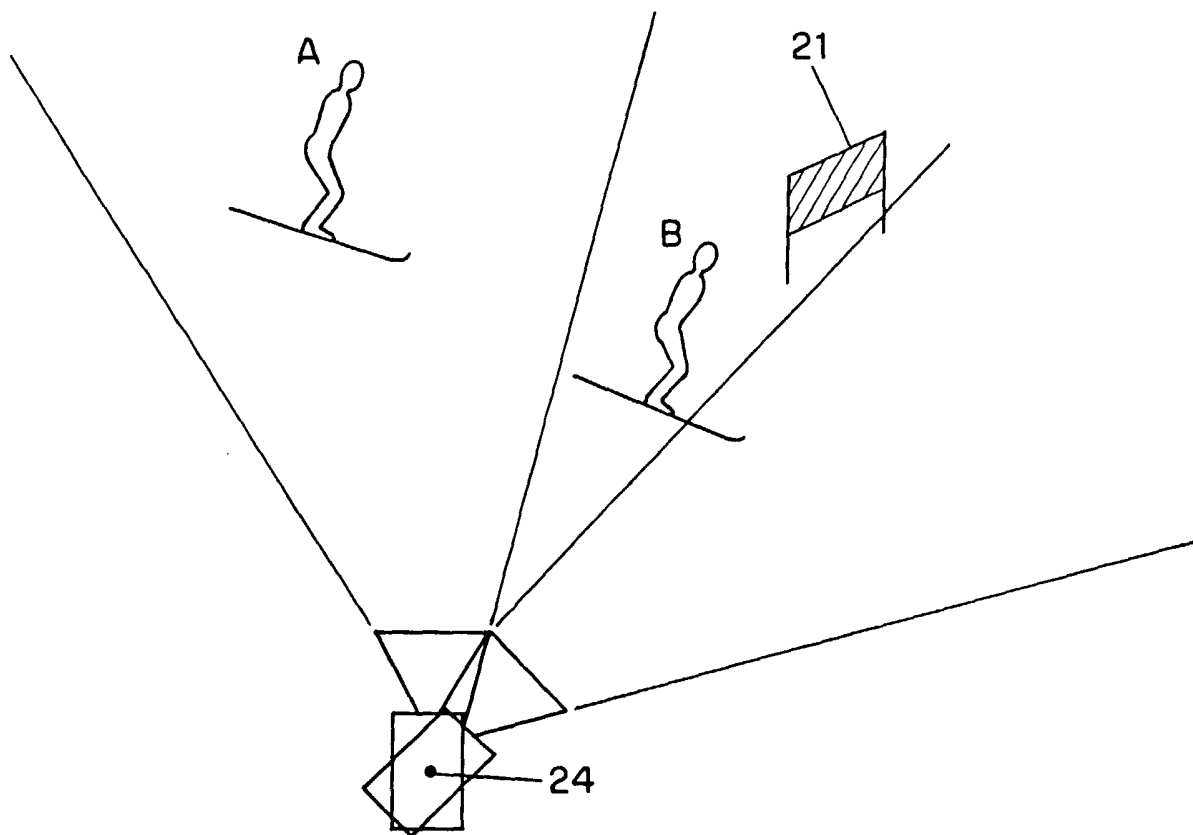


FIG. 4

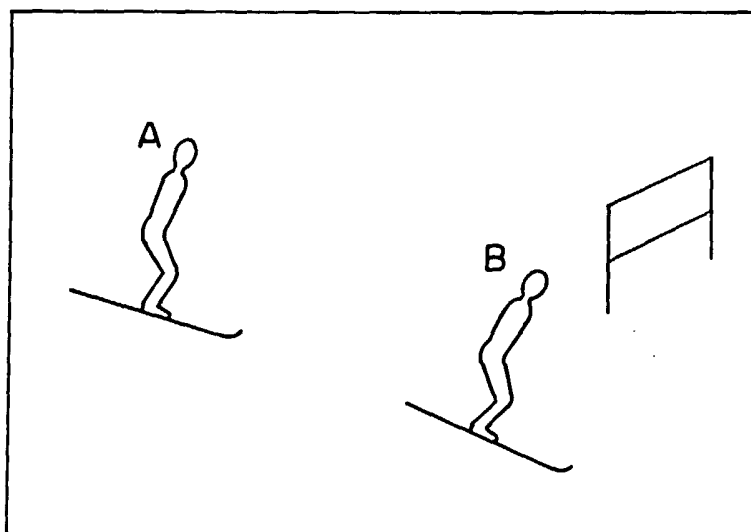


FIG. 5

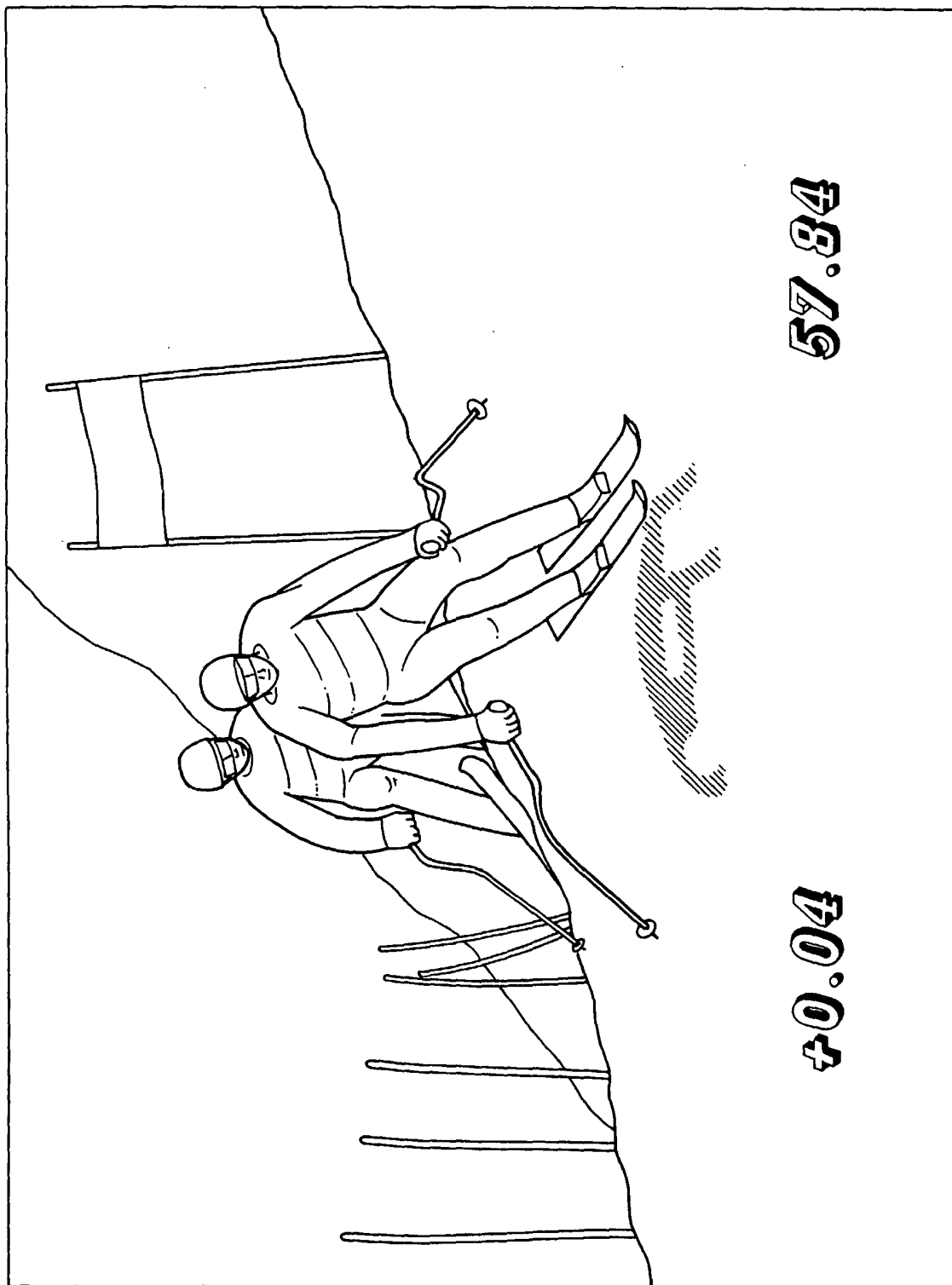


FIG. 6

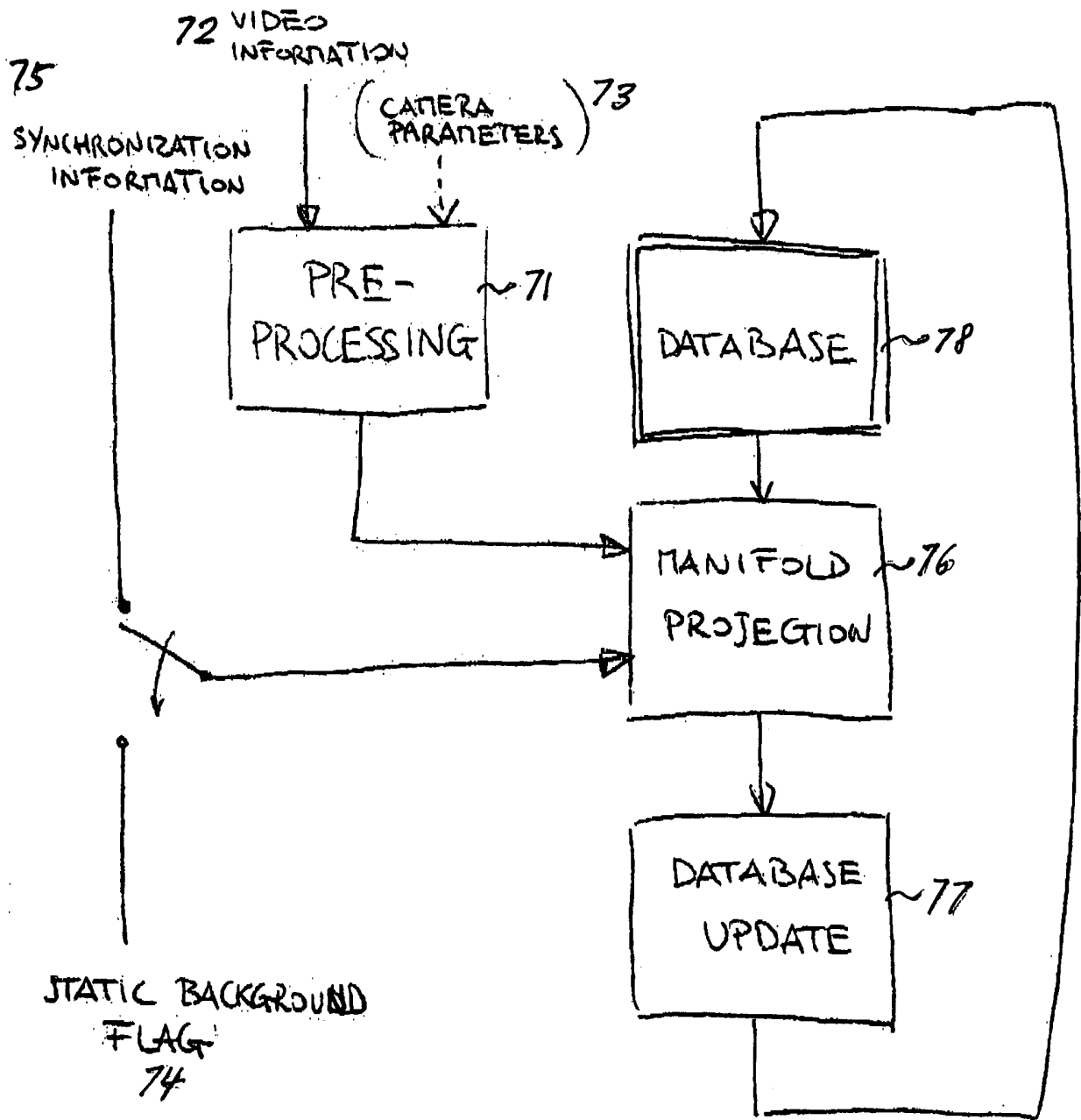


Fig. 7

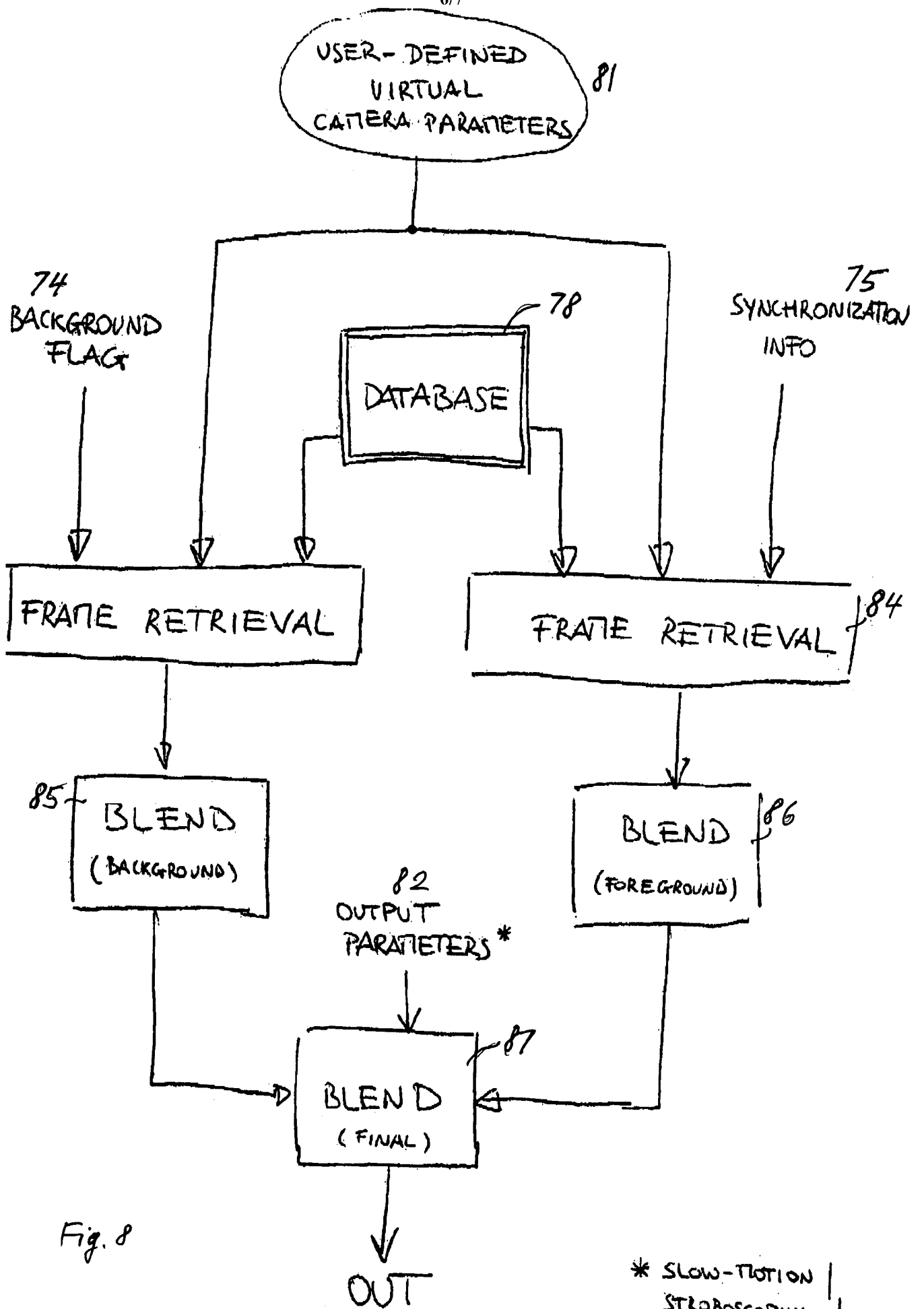


Fig. 8

\* SLOW-MOTION / STROBOSCOPING

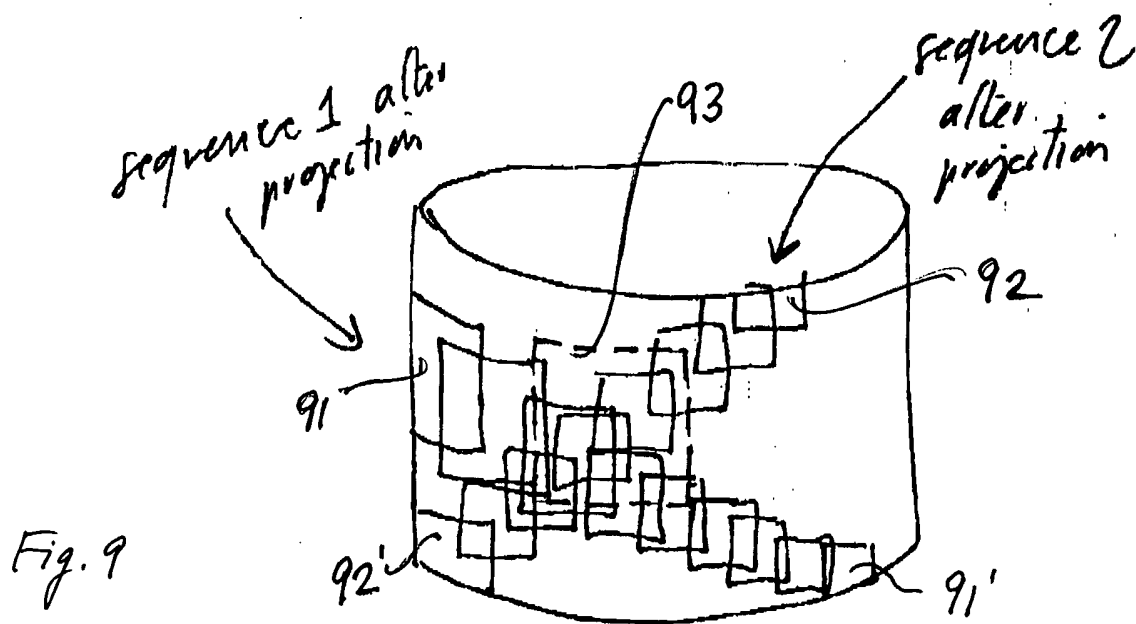


Fig. 9

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US01/19741

**A. CLASSIFICATION OF SUBJECT MATTER**  
 IPC(7) :H04N 5/272  
 US CL : 348/584  
 According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**  
 Minimum documentation searched (classification system followed by classification symbols)  
 U.S. : 348/584, 585, 586, 588, 578, 571; H04N 5/265, 5/272, 5/262

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
 none

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
 none

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 4,843,483 A (BOGNER) 27 JUNE 1989 see entire document	1-43
A	DE 4,135,385 A1 (BAUER) 14 MAY 1992 see abstract	1-43
A	WO 90/04848 (STRONG) 03 MAY 1990 see abstract	1-43

Further documents are listed in the continuation of Box C.  See patent family annex.

"A"	document defining the general state of the art which is not considered to be of particular relevance	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"E"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means	"&"	document member of the same patent family
"I"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 01 SEPTEMBER 2001	Date of mailing of the international search report <b>15 MAR 2002</b>
--	--

Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized Officer <i>David E. Harvey</i> DAVID E. HARVEY Telephone No. (703) 305-4865
--	---