

A SYSTEM FOR IMPLANTING AN IMAGE INTO A VIDEO STREAM

5

10

FIELD OF THE INVENTION

The present invention relates generally to merging of a prepared image with a video signal.

15

BACKGROUND OF THE INVENTION

Sports arenas typically include a game area where the game occurs, a seating area where the spectators sit and a wall of some kind separating the two areas. Typically, the wall is at least partially covered with advertisements from the companies which sponsor the game. When the game is filmed, the advertisements on the wall are filmed as part of the sports arena. The advertisements cannot be presented to the public at large unless they are filmed by the television cameras.

Systems are known which merge predefined advertisements onto surfaces in a video of a sports arena. One system has an operator define a target surface in the arena. The system then locks on the target surface and merges a predetermined advertisement with the portion of the video stream corresponding to the surface. When the camera ceases to look at the surface, the system loses the target surface and the operator has to indicate again which surface is to be utilized.

The above-described system operates in real-time. Other systems are known which perform essentially the same operation but not in real-time.

Other systems for merging data onto a video sequence are known. These include inserting an image between video scenes, superposition of image data at a fixed

location of the television frame (such as of television station logos) and even electronic insertion of image data as a "replacement" of a specific targeted billboard. The latter is performed using techniques such as color
5 keying.

U.S. 5,264,933 describes an apparatus and method of altering video images to enable the addition of advertising images to be part of the image originally displayed. The operator selects where in the captured
10 image the advertising image is to be implanted. The system of U.S. 5,264,933 can also implant images, in selected main broadcasting areas, in response to audio signals, such as typical expressions of commentators.

PCT Application PCT/FR91/00296 describes a procedure
15 and device for modifying a zone in successive images. The images show a non-deformable target zone which has register marks nearby. The system searches for the register marks and uses them to determine the location of the zone. A previously prepared image can then be
20 superimposed on the zone. The register marks are any easily identifiable marks (such as crosses or other "graphemes") within or near the target zone. The system of PCT/FR91/00296 produces the captured image at many resolutions and utilizes the many resolutions in its
25 identification process.

SUMMARY OF THE PRESENT INVENTION

It is an object of the present invention to provide a system and method which mix images, such as an
30 advertisement, with a video stream of action occurring within a relatively unchanging space. Such a space may be a playing field or court, a stage or a room and the location is typically selected prior to the action (e.g. game or show). The images are "implanted" onto a
35 selected surface of the background space, where the term "implanted" herein means that the images are mixed onto

the part of the video stream showing the selected surface.

Specifically, the present invention utilizes apriori information regarding the background space to
5 change the video stream so as to include the image at some location within the background space. The system and method operate no matter which perspective view of the background space is presented in the video stream.

In accordance with a preferred embodiment of
10 the present invention, the system preferably includes a video frame grabber and an image implantation system. The frame grabber grabs a single frame of the video signal at one time. The image implantation system typically implants the advertising image into the frame
15 onto a predefined portion of a preselected one of the surfaces of the background space if the portion is shown in the frame. To determine the location of the portion to receive the implantation, the image implantation system includes a unit for receiving a) a flat model of the
20 fixed surfaces of the background space and b) an image mask indicating the portion of the flat model onto which the image is to be mixed. Via the model, the image implantation system identifies if and where the portion is shown in the frame.

25 Moreover, in accordance with a preferred embodiment of the present invention, the system also includes a design workstation on which the image and an image mask which indicates the preselected surface can be designed.

30 Further, the identification preferably involves a) reviewing the frame and extracting features of the fixed surfaces therefrom and b) determining a perspective transformation between the model and the extracted features.

35 Still further, the reviewing and extracting includes creating a background mask and a foreground

mask. The background mask indicates the locations of features of interest, of background elements in the frame and is utilized to extract desired features. The foreground mask is formed of the foreground elements of the frame which must remain unchanged.

Additionally, in accordance with a preferred embodiment of the present invention, the implantation includes the steps of a) transforming the image, an image mask and, optionally, a blending mask, with the perspective transformation, and b) mixing the transformed image, image mask and optional blending mask with the frame and with the foreground mask. The foreground mask, as mentioned hereinabove, indicates locations of foreground data not to be covered by the transformed image.

Further, the system preferably includes a lookup table for converting between the multiplicity of colors in the frame to one of: colors of features of interest, colors of background elements and a color indicating foreground elements. The lookup table is preferably created by having a user indicate the relevant colors. If the relevant colors no longer indicate the features of interest and the background elements (typically due to lighting changes), the user can indicate new colors which do indicate the desired elements and the lookup table is then corrected.

Still further, in accordance with a preferred embodiment of the present invention, the lookup table is utilized to create the background and foreground masks of the frame indicating the locations of features of interest, of background elements and of foreground elements in the frame.

In accordance with an exemplary embodiment of the present invention, the features are lines. In one embodiment, they are extracted with a Hough transform. In another embodiment, they are extracted by determining

the angles of line segments. Pixels of interest are selected and a neighborhood opened. The neighborhood is subdivided and the sector having the greatest activity is selected. The selected sector is then extended and
5 divided. The process is repeated as necessary.

Moreover, in accordance with a preferred embodiment of the present invention, the system projects the extracted features onto an asymptotic function to determine which of the features are perspective versions
10 of parallel lines.

Further, in accordance with the exemplary embodiment of the present invention, the background space is a sports arena having lines marked on it. The system has a model of the sports arena and, preferably, has a
15 list of the rectangles in the model and the locations of their corner points. The system preferably performs the following operations:

a) selects two vertical and two horizontal lines from the extracted features and determines their
20 intersection points;

b) generates a transformation matrix from the corner points of each rectangle of the model to the feature intersection points;

c) transforms the model with each
25 transformation matrix;

d) utilizing the background elements of the background mask, matches each transformed model with the frame; and

e) selects the transformation matrix which
30 matches the features of the frame best.

Moreover, in accordance with the exemplary embodiment of the present invention, camera parameters can be utilized to reduce the number of lines in the frame needed to identify the sports field. For this
35 embodiment, the following actions occur:

receiving or extracting the coordinates of a

set of cameras;

representing a current transformation matrix as
a product of coordinate, tilt, turn and zoom matrices and
then determining the values for the tilt, turn and zoom;
5 and

identifying the camera having the calculated
values for tilt, turn and zoom and storing the
information; and

repeating the steps of receiving, representing
10 and identifying whenever there is a new cut in the video.

Any frame in the video stream can now be
treated as either being similar to the previous frame or
as part of a new cut taken by an identified camera.

15

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be understood and
appreciated more fully from the following detailed
description taken in conjunction with the drawings in
20 which:

Fig. 1 is a block diagram illustration of a
system for implanting images into a video stream,
constructed and operative in accordance with a preferred
embodiment of the present invention;

25 Fig. 2 is a schematic illustration of a tennis
game used as an example for explaining the operation of
the system of Fig. 1;

Fig. 3 is an illustration of a model of a
tennis court, useful in understanding the operation of
30 the system of Fig. 1;

Fig. 4A is an illustration of an image to be
implanted;

Fig. 4B is an illustration of an image region
mask for the image of Fig. 4A and the model of Fig. 3;

35 Fig. 4C is an illustration of a blending mask
for the image of Fig. 4A and the model of Fig. 3;

Fig. 5 is a block diagram illustration of elements of an image implantation unit forming part of the system of Fig. 1;

Fig. 6 is an illustration of an exemplary video frame into which the image of Fig. 4A is to be implanted;

Fig. 7 is an illustration of a background mask generated from the video frame of Fig. 6;

Fig. 8 is a block diagram illustration of the operations of a feature identification unit forming part of the image implantation unit of Fig. 5;

Fig. 9A is a flow chart illustration of a method of feature extraction;

Fig. 9B is an illustration of a portion of the background mask, useful in understanding the method of Fig. 9A;

Fig. 9C is an illustration of a histogram of subsectors of the background mask of Fig. 9B, useful in understanding the method of Fig. 9A;

Fig. 10 is a block diagram illustration of the operations of a perspective identification unit forming part of the image implantation unit of Fig. 5;

Fig. 11A is an illustration of the meeting points of extracted features from Fig. 7;

Fig. 11B is an illustration of perspective parallel lines meeting at different points due to calculation inaccuracies;

Figs. 12A and 12B are illustrations of gnomonic projections, useful in understanding the operations of the perspective identification unit of Fig. 10;

Fig. 12C is a graphical illustration of an exemplary function useful for the gnomonic projections of Figs. 12A and 12B;

Fig. 13 is a detailed block diagram illustration of the operations illustrated in Fig. 10;

Figs. 14A and 14B are useful in understanding the operations of Fig. 13;

Fig. 15 is an illustration of the use of transformation matrices;

Fig. 16 is an illustration useful in understanding the matching process between quadrilaterals and the geometric model, useful in understanding the operations of Fig. 13;

Fig. 17 is a block diagram illustration of the operations of transformer and mixing units of the image implantation unit of Fig. 5;

Fig. 18 is a block diagram illustration of a correction method for updating a lookup table used in the image implantation unit of Fig. 5;

Fig. 19 is a schematic illustration of camera parameters;

Fig. 20 is a flow chart illustration of transformation matrix operations when the camera parameters of Fig. 19 are known or calculable;

Fig. 21 is an illustration of a table useful in the process shown in Fig 20; and

Fig. 22 is a flow-chart illustration of a method of operation when the camera parameters are known or calculable.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Reference is now made to Fig. 1 which illustrates a system for mixing images, such as advertisements, with a video stream of action occurring within a relatively unchanging background space. The images are implanted onto a selected surface of the background space. The system will be described in the context of a video of a tennis game, illustrated in Fig. 2 to which reference is also made. It will be understood that the present invention is operative for all situations in which the surfaces in which action occurs are known a priori and are identifiable.

The system of the present invention typically

comprises a video frame grabber 10 for converting an input video sequence (such as of a tennis game) into video frames, a design workstation 12 for designing images (such as advertisements) to be implanted onto a selected surface (such as on the tennis court) seen within the video frame, an image implantation system 14 for merging the video frame with the designed image, a control computer system 16 for controlling the action of and providing operator input to the image implantation system 14 and a broadcast monitor 18.

The control computer system 16 typically comprises a central processing unit (CPU) 20, a keyboard 22, a mouse 24, a disk 26, a removable media drive such as a floppy 27, and a monitor 28. The monitor 28 is typically driven by a graphics adaptor forming part of the CPU 20. The design workstation 12 typically also includes a removable media drive such as floppy 27.

The control computer system 16 and the image implantation system 14 typically communicate via a system bus 29. The design workstation and the control computer system 16 typically communicate via removable media.

The video sequence can be received from any source, such as a videotape, a remote transmitting station via satellite, microwave or any other type of video communication, etc. If the sequence is provided from a satellite, the system has no control over the video rate. Therefore, the image implantation system 14 must perform its operations within the video rate, typically 30ms between frames, of the satellite video stream. If the sequence comes from a videotape, the system can control the video rate and operate at any desired speed.

The video sequence is originally produced at the site of the game. As can be seen in Fig. 2, for tennis games, there are typically two television cameras viewing the action on the tennis court 32. The

locations of the television cameras 30 typically are fixed.

5 The court 32 is divided into two halves by a net 34. Each half has a plurality of areas 36, typically painted a first shade of green, divided by a plurality of lines 38, typically painted white. The outer court area 40 is typically painted a second shade of green.

10 In reality, the lines 38 are parallel and perpendicular lines. Since the cameras 30 zoom in on the action from an angle, rather than from above, the images of the action which they receive are perspective views. Thus, in the video output of the cameras 30, the parallel lines 38 appear as though they converge at infinity. The angle of perspective of the video output changes as the
15 angles of the cameras 30 and the amount of zoom.

The present invention will implant an image 42, such as the word "IMAGE", at a desired location on a selected background surface, for all perspective angles and amount of zoom. For tennis courts, the possible
20 locations are any rectangles within one half of the tennis court 32 defined by four lines 38. As shown in Fig. 2, the image 42 will not interfere with the action of players 44; it will appear as though the image 42 was painted on the surface of the court.

25 Since, in reality, the shape of court 32 and the location of lines 38 within the court 32 do not change, if the image implantation system has a model of the playing space, including the location in which the image is to be implanted, and can identify at least the
30 viewing angle and amount of zoom, it can merge the image into the video sequence so that it will appear as though the image was implanted at the desired location. To do this, the image implantation system additionally needs to know the colors of the court as seen by the cameras.
35 These colors can change as the lighting (daylight or artificial) change.

Reference is now additionally made to Fig. 3 which illustrates a geometric model 50 of the tennis court and to Figs. 4A, 4B and 4C which illustrate data which an implantation designer prepares.

5 The implantation designer works at the design workstation 12, such as the BLAZE workstation manufactured by Scitex Corporation Ltd. of Herzlia, Israel, and typically has the geometric model 50 of the tennis court 32, typically as a top view. The model 50
10 is typically a scaled version of the court 32, indicating the elements of it which are to be identified by the implantation system 14, such as the lines 38. Other playing fields may include circles or other well-defined curves. Other identifiable elements include
15 intersections 54 of the lines 38.

 The implantation designer designs the image 42 (illustrated in Fig. 4A) to be implanted and determines where on the model 50 to place it. A number of possible locations 52 are shown in Fig. 3. The designer then
20 prepares an image location mask 56 (Fig. 4B) to identify where within the model 50 the image 42 is to be placed. The mask 56 is light at the location in model 50 where the image 42 is to be placed and dark everywhere else.

 Since the image 42 may be of bright colors, it
25 may be desired not to implant the image itself but a softened version of it, so as not to significantly disturb the action on the court 32. Therefore, the implantation designer may optionally prepare a blending mask 58 (Fig. 4C) indicating how the image 42 is to be
30 blended with the color of the court 32 at the location of implantation as indicated by location mask 56. The blending mask 58 can be any suitable mask such as are known in the art. In Fig. 4C, mask 58 is shown to have four areas 59, each indicating the inclusion of a
35 different amount of court color, where the outer area 59 typically incorporates much more of the court color than

the inner areas.

Reference is now made back to Figs. 1 and 2. The implantation data, formed of the geometric model 50, the image 42, the image location mask 56 and the optional
5 blending mask 58, are typically prepared before the relevant tennis match and are provided to the image implantation system 14, typically via removable media, for implantation into the input video sequence when the match occurs.

10 Most video sequences of live televised games begin with an initializing sequence operative to enable local station operators to synchronize their systems to the input sequence. This is also typically true for taped video data.

15 In the present invention, the initializing video data is grabbed by the frame grabber 10 and is provided first to the control computer system 16. A station operator selects a frame which has a clear view of the game field and uses it to provide calibration
20 information, as described hereinbelow. The calibration information is utilized by the image implantation system 14 to identify the court 32 and its features (such as lines 38). In the embodiment described hereinbelow, the calibration information includes the colors of the
25 features of interest in the background, such as the field lines, the playing field (court 32) and the ground outside the playing field (outer court area 40). The remaining colors which may be received are defined as foreground colors. Other playing fields may require
30 fewer or more features to define them and thus, fewer or more colors.

The station operator, utilizing the mouse 24 and keyboard 22, interactively defines the calibration colors. This can be achieved in a number of ways, one of
35 which will be described herein. A four color layer is superimposed over the frame currently displayed on

control monitor 28. Initially, the four color layer is comprised of one color only, a transparent color. Thus, the current frame is initially visible.

The operator indicates pixels describing one of the three features, lines 38, inner playing field 36 and outer playing field 40. When he selects a pixel, those pixels in the superimposed layer which correspond to pixels in the current frame having the selected color are colored a single translated color, thereby covering their corresponding pixels of the current frame. The selected color is stored. The process is repeated for all three areas. All colors not selected are assigned a fourth translated color.

If the operator approves the resultant four color layer, a lookup table (LUT) between the colors selected from the current frame and the translated colors is produced.

If desired, the control computer system 16 can store the pixels which the operator selected for later use in a LUT correction cycle, described hereinbelow with reference to Fig. 18.

The control computer system 16 provides the frame data, consisting of the LUT and the pixels utilized to produce the LUT, to the image implantation system 14. System 14 utilizes the above described frame data to identify the desired features in each frame of the input video signal.

Reference is now made to Fig. 5 which illustrates the general elements of the image implantation system 14. Reference is also made to Figs. 6 and 7 which are useful in understanding the operation of the system 14.

The system 14 typically comprises a feature identification unit 60 (Fig. 5) for identifying which features of the court 32 are present in each input video frame and a perspective identification unit 62 for

identifying the viewing angle and zoom of an active camera 30 and for determining an appropriate perspective transformation between the model 50 and the input video frame. The system 14 also comprises a transformer 64 for
5 transforming the implantation data from the model plane to the image viewing plane and a mixer 66 for mixing the perspective implantation data with the current video frame, thereby to implant the image 42 onto the court 32.

As described in more detail hereinbelow, the
10 feature identification unit 60 utilizes the LUT to create a background mask of the input frame indicating which parts of the frame have possible background features of interest and which parts are foreground and therefore, are not to be changed in later operations. Figs. 6 and 7
15 respectively provide an exemplary input frame 68 and its corresponding background mask 70.

The input frame 68 of Fig. 6 has two players 44 on the court 32. The background mask 70 of Fig. 7 shows the areas of the four colors. The areas marked 1 - 4 are
20 the areas of line color, inner court color, outer court color and remaining colors, respectively. It is noted that the areas of the players 44 are marked with the background color 4 and cover over other important areas, such as those of the white lines 1.

25 From the background mask 70, unit 60 (Fig. 5) extracts the features of the playing field. For tennis courts, the features of interest are the lines 38. The perspective identification unit 62 compares the extracted features with those of the model 50 and produces
30 therefrom a transformation matrix.

Using the transformation matrix, the transformer 64 converts the image implantation data (i.e. image 42 to be implanted, the image location mask 56 and the blending mask 58) to the perspective of the input
35 video frame.

Finally, using the transformed image location

mask 56 and the background mask 70, the mixer 66 implants the perspective version of image 42 into the desired background parts of the input video frame. Thus, if the players walk on the part of the court 32 where the image 42 is implanted, they will appear to walk "over" the implanted image. If desired, the transformed blending mask 58 can be utilized to blend the image 42 with the colors of the field on which the image 42 is implanted.

Reference is now made to Fig. 8 which details the operations of the feature identification unit 60. In step 72, unit 60 uses the LUT to convert the input video frame from a many colored frame to the four color picture called the background mask 70. Specifically, for the tennis court 32, the LUT provides a first value to pixels having colors of the lines 38, a second value to pixels having colors of the inner court 36, a third value to pixels having colors of the outer court 40 and a fourth value (indicating foreground pixels) to the remaining pixels. This is shown in Fig. 7. The LUT can be implemented in any suitable one of the many methods known in the art.

The background mask 70 not only defines which pixels belong to the background of interest, it also includes in it the features of interest, such as lines 38. Thus, in step 74, the feature identification unit 60 processes background mask 70 to extract the features of interest. Typically though not necessarily, the LUT is designed to provide the features with a single color value.

For the example of a tennis match, the extraction involves reviewing those pixels of the background mask 70 having the first value and extracting straight segments therefrom. For example, step 74 can be implemented with a Hough transform operating on the background mask 70. Hough transforms are described on pages 121 - 126 of the book Digital Picture Processing,

Second Edition, Vol. 2 by Azriel Rosenfeld and Avinash C. Kak, Academic Press, 1982, which book is incorporated herein by reference.

5 The result is an array of line parameters, each describing one straight segment in the background mask 70. The line parameters for each segment include the coefficients of the line equations describing it as well as a weight value indicating the number of pixels included within the segment.

10 An alternative method of extraction is illustrated in Figs. 9A, 9B and 9C to which reference is now briefly made. As shown generally in Fig. 9A, the method begins at a first pixel 69 (Fig. 9B) of the background mask 70 having the color of interest (in this
15 example, white) and looks in its neighborhood 75 to determine where there are more white pixels (marked by shading). To do so, it divides the neighborhood 75 into subsectors 71 - 74 of a predetermined size and performs a histogram of distribution of white pixels in the each
20 subsector. Fig. 9C illustrates the histogram for the sectors 71 - 74 of Fig. 9B. The one with a strong maximum (subsector 73) is selected as the next sector for searching.

25 In the next step, a new neighborhood 78 is defined which consists of the selected subsector 73 and an extension thereof. The entire neighborhood 78 is twice as long as the neighborhood 75. This new neighborhood 78 is subdivided into four subsectors 76 and the process repeated.

30 This process continues until one of the following criteria are met:

1. the sub-sector is narrow enough to be defined as a straight line;
2. no strong maximum is obtained in the
35 histogram.

 If condition 1 is obtained, the coefficients of

the straight line are stored and the pixels forming the straight line are then "colored" to have the "remaining color" and so eliminated from the search.

5 The feature extraction process produces an array of possible features which includes the true features as well as stray lines.

Reference is now made to Fig. 10 which illustrates, in general, the operations of the perspective identification unit 62 of Fig. 5. Reference
10 is also made to Figs. 11A and 11B which are useful in understanding the operation of unit 62 in general, to Fig. 13 which details the operations of unit 62 for the example of the tennis court 32 and to Figs. 12A, 12B, 12C, 14A and 14B which are useful in understanding the
15 operations detailed in Fig. 13.

Using a priori information, unit 62, in step 80, processes the array of possible features and determines which ones are most likely to be the features of interest. In step 82, unit 62 selects a minimum set
20 of features from the resultant true features and attempts to match them to features of the model 50. The process is repeated as often as necessary until a match is found. In step 84, the matched features are utilized to generate a transformation matrix M transforming the model to the
25 features in the input video frame.

In the example of the tennis court 32, step 80 utilizes the fact that the lines 38 of model 50 are parallel in two directions (vertical and horizontal) and that in perspective views (such as in the input video
30 frame), lines which are parallel in reality meet at a finite point. This is illustrated in Fig. 11A in which all the extracted line segments, represented by solid lines, are extended by dashed lines. The perspective lines which correspond to parallel lines in reality (e.g.
35 pseudo parallel lines 90) intersect at a point 91 far from the outer edges 92 of the frame. All other

intersections, labeled 94, occur within the edges 92 or close to its borders.

However, as illustrated in Fig. 11B, because of digitization errors, it might be determined that the extensions of three pseudo parallel lines do not meet at a single point. In fact, they might meet at three widely separated points 96.

Applicants have realized that, since perspective parallel lines do meet at infinity, the projection of the extracted lines onto an asymptotic function will cause the intersection points to occur close together. Therefore, in accordance with a preferred embodiment of the present invention, the extracted line segments are projected onto a two-dimensional asymptotic function. One such projection is known as a "Gnomonic Projection" and is described on pages 258, 259 and 275 of the book Robot Vision by Berthold Klaus Paul Horn, The MIT Press, Cambridge, Massachusetts, 1986, which pages are incorporated herein by reference. Examples of gnomonic projections are illustrated in Figs. 12A and 12B.

In the gnomonic projection, a point 100 on an XY plane 102 is projected onto a point 100' on a hemisphere 104. A line 106 in the XY plane is projected onto a great arc 106' of the hemisphere 104 (i.e. an arc of a great circle of a sphere). The origin is represented by the south pole 109 and infinity is represented by the equator 108. Thus, any cluster 110 (Fig. 12B) of points near the equator 108 represents the intersection of pseudo parallel lines and thus, the lines which have points going through a cluster 110 are parallel lines.

Fig. 12B illustrates a plurality of great arcs, labeled 120a - 120f, corresponding to some arbitrary extracted line segments (not shown). The three arcs 120a - 120c have intersection points 122 which form a cluster

110a near the equator 108. Great arcs 120d - 120f also intersect near the equator, but at cluster 110b. All of the great arcs intersect each other, but their other intersections are at locations closer to the south pole
5 109 than to the equator 108.

In step 130 (Fig. 13), the gnomonic projection is utilized to produce an array of great arcs from the array of straight line segments produced from the feature extraction (step 74, Fig. 8).

10 In step 132, the area around equator 108 is searched to find all of the intersection points 122. A value V_k is given to each intersection point. The value V_k is a function of the weights W_i of the line segments which intersect and the Z coordinate of the intersection
15 point 122. An example of a function V_k is provided in equation 1:

$$V_k = W_{\text{line 1}} * W_{\text{line 2}} * f(Z_{\text{intersection point}}) \quad (1)$$

20 where $f(Z_{\text{intersection point}})$ is any function having a curve similar to curve 134 of Fig. 12C wherein most points receive a low value and only those points approaching the equator 108 ($Z=1$) receive values close to 1. For example, $f(Z_{\text{intersection point}})$ might be Z^5 .

25 In step 136, a small neighborhood around each intersection point 122 is searched for other intersection points. If any are found, the present intersection point and the ones found are stored as a cluster 110 (Fig. 12B). A cluster 110 is also defined as one whose value
30 of $f(Z_{\text{intersection point}})$ is above a predefined threshold. Thus, a cluster 110 can include only one intersection point. In Fig. 12B there are three clusters 110a - 110c, one of which, cluster 110c, includes only one intersection point 122.

35 Once all of the points have been searched, a location of each cluster 110 is determined by finding the

"center of gravity" of the points in the cluster. The weight of the cluster 110 is the sum of the values V_k of the points in the cluster.

5 In step 138, the two clusters with the highest weights are selected. For the example of Fig. 12B, clusters 110a and 110b are selected.

10 In step 140, one cluster is assumed to represent "vertical" lines and the other to represent "horizontal" lines. Also, in step 140, the straight segments corresponding to the lines of the two selected clusters are marked "vertical" or "horizontal", respectively.

15 In step 142, the "vertical" and "horizontal" lines are reviewed and the two heaviest vertical and two heaviest horizontal lines are selected, where "heaviest" is determined by the values of W_i . The selected lines, labeled 146, are shown in Fig. 14A for the lines of Fig. 11A. In step 144 the intersection points, labeled A, B, C and D, of the four selected lines are determined and stored. As shown in Fig. 14A, the selected lines may intersect out of the frame.

20 Steps 130 - 144 are the operations needed to identify the true features in the video frame (step 80 of Fig. 10). The output of step 144 are the features which are to be matched to the model. The remaining steps match the features to the model and determine the transformation (steps 82 and 84 of Fig. 10) as an integrated set of operations.

30 A standard tennis court has five vertical lines and four horizontal lines. Since it is not possible to differentiate between the two halves of the court, only three horizontal lines are important. The number of different quadrilaterals that can be formed from a selection of two horizontal lines out of three (three possible combinations) and two verticals out of five (10 possible combinations) is thirty. The thirty

quadrilaterals may be in four different orientations for a total of 120 rectangles.

In step 150, one of the 120 rectangles in the geometric model 50 is selected by selecting its four corners, labeled A', B', C' and D' (Fig. 14B). As can be seen, this is not the correct match.

In step 152, the matrix M, which transforms from the four points A', B', C' and D' of the model (Fig. 14B) to the four points A, B, C, D of the video frame (Fig. 14A), is determined. The matrix M can be represented as a superposition of subsequent transformations as explained with reference to Fig. 15.

Fig. 15 shows three quadrilaterals 180, 182 and 184. Quadrilateral 180 is the model quadrilateral ABCD shown in an XY plane, quadrilateral 182 is a unit square having points (0,1), (1,1), (0,0) and (1,0) in a TS plane, and quadrilateral 184 is the perspective quadrilateral 184 in a UV plane.

The transformation M from model quadrilateral 180 to perspective quadrilateral 184 can be represented by the superposition of two transformations, a translation and scaling matrix T from quadrilateral 180 to the unit square 182 and a perspective matrix P from the unit square 182 to quadrilateral 184. Matrix T, in homogeneous coordinates, has the form:

$$T = \begin{vmatrix} Sx & 0 & 0 \\ 0 & Sy & 0 \\ Tx & Ty & 1 \end{vmatrix} \quad (2)$$

where Sx and Sy are the scaling factors in the X and Y directions, respectively and Tx and Ty are the X and Y translation factors. Sx, Sy, Tx and Ty are determined by the equation:

$$(x,y,1)*T = (s,t,1) \quad (3)$$

for the four coordinates (x,y,1) of quadrilateral 180 and

the four coordinates (s,t,1) of unit square 182.

Matrix P, in homogeneous coordinates, has the form:

$$P = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \quad (4)$$

The elements of the matrix P are determined by solving the following equation:

$$(s,t,1)*P = (u,v,w) \quad (5)$$

where (u, v, w) represents the four known coordinates of the points A, B, C and D of quadrilateral 184, as shown in Fig. 15, and w is always normalized.

Assume $a_{33} = 1$, then P can be calculated as follows:

From (s,t,1) = (0,0,1), we determine that

$$\begin{aligned} a_{31} &= U_{00} \\ a_{32} &= V_{00} \end{aligned} \quad (6)$$

From (s,t,1) = (1,0,1), we determine that:

$$\begin{aligned} a_{11} + a_{31} &= U_{10}(a_{13} + 1) \Rightarrow a_{11} = U_{10}(a_{13} + 1) - U_{00} \\ a_{12} + a_{32} &= V_{10}(a_{13} + 1) \Rightarrow a_{12} = V_{10}(a_{13} + 1) - V_{00} \end{aligned} \quad (7)$$

From (s,t,1) = (0,1,1) we determine that:

$$\begin{aligned} a_{21} + a_{31} &= U_{01}(a_{23} + 1) \Rightarrow a_{21} = U_{01}(a_{23} + 1) - U_{00} \\ a_{22} + a_{32} &= V_{01}(a_{23} + 1) \Rightarrow a_{22} = V_{01}(a_{23} + 1) - V_{00} \end{aligned} \quad (8)$$

From (s,t,1) = (1,1,1) we determine that:

$$\begin{aligned} a_{11} + a_{21} + a_{31} &= U_{11}(a_{13} + a_{23} + 1) \\ a_{12} + a_{22} + a_{32} &= V_{11}(a_{13} + a_{23} + 1) \end{aligned} \quad (9)$$

From equations 7 - 9, two equations in two unknowns, a_{13} and a_{23} , are produced, as follows:

$$\begin{aligned} a_{13}(U_{10} - U_{11}) + a_{23}(U_{01} - U_{11}) &= U_{11} + U_{00} - U_{10} - U_{01} \\ a_{13}(V_{10} - V_{11}) + a_{23}(V_{01} - V_{11}) &= V_{11} + V_{00} - V_{10} - V_{01} \end{aligned} \quad (10)$$

Once a_{13} and a_{23} are determined, the remaining elements can be determined from equations 7 and 8.

The transformation, or mapping, matrix M is the matrix product of matrices T and P , as follows:

5

$$M = T * P \quad (11)$$

In step 154, the lines 38 of the model 50 are mapped onto the video frame using the mapping matrix M .

10 The result is a distorted frame 156 (Fig. 16) having 1s wherever the converted pixels of the model exist and 0s everywhere else. As can be seen, the points A' , B' , C' and D' match the points A , B , C and D , respectively. However, the rest of the geometric model 50 does not.

15 In step 158 the distorted frame 156 is XOR'd with the background mask 70 (Fig. 7). The XOR step outputs a 0 on two occasions: a) the pixels of the distorted frame 156 have a 1 value and the pixels of the video frame have the field line color; and b) the pixels of the distorted
20 frame 156 have a 0 value and the pixels of the video frame have the "non-line" color. The remaining situations receive 1 values.

In steps 160 and 161 the number of pixels having 1 values are counted and the value is associated with
25 transformation matrix M .

After all of the matrices M have been determined, in step 162 the matrix having the least weight is selected. Since there is a possibility that no match can be made (i.e. the video is showing a commercial, the
30 television cameras 30 are viewing the audience, etc.), in step 164, the weight of the selected matrix is checked against a threshold. If it is above that value, then a null transformation matrix is provided. Otherwise, the selected matrix is defined as the transformation matrix
35 M . Null transformation matrices are also provided when test conditions, of any of the previous steps, fail.

Reference is now made to Fig. 17 which illustrates the operations of the transformer 64 and the mixer 66 of Fig. 5. The transformer 64 utilizes the transformation matrix M to distort each of the image 42, the image region mask 56 and the blending mask 58 into the plane of the video frame (step 170). It also ANDs the distorted image region mask with the background mask 70, producing therefrom a permission mask. The permission mask indicates those pixels of the video frame which are both background pixels and within the image region. Onto these pixels the image will be implanted.

The mixer 66 combines the distorted image with the video frame in accordance with the blending and permission masks. The formula which is implemented for each pixel (x,y) is typically:

$$\text{Output}(x,y) = \beta(x,y) * \text{image}(x,y) + (1 - \beta(x,y)) * \text{video}(x,y) \quad (12)$$

$$\beta(x,y) = \alpha(x,y) * P(x,y) \quad (13)$$

where Output(x,y) is the value of the pixel of the output frame, image(x,y) and video(x,y) are the values in the implanted image 42 and the video frame, respectively, $\alpha(x,y)$ is the value in the blending mask 58 and P(x,y) is the value in the permission mask.

The description hereinabove assumes that the LUT which produced the background mask 70 remains correct during the entire game. If the lighting changes (which typically occurs in outdoor games), the colors in the video sequence can change and, as a result, the background mask 70 will no longer correctly indicate the background elements. Therefore, a correction procedure can be periodically performed. The correction procedure is detailed in Fig. 18 to which reference is now made.

It is noted that, in the calibration process, test spots, indicating the features of interest in the background (such as field lines and the inner and outer

courts) were selected by the operator. The spot locations were saved, as were their color values.

Once the matrix for the calibration video frame is determined, the locations of the test spots are converted from the video frame plane into the geometric model plane (i.e. by using the inverse of the matrix M). At some later time when calibration is desired, the test spots are converted into the current video frame plane. The distorted test spots which are within the current video frame are selected and their neighborhoods sampled. The color characteristics of each neighborhood are calculated (using histograms, for example) and the result is compared to the characteristics of the saved spot. If there is any significant change in the colors, the LUT is corrected and the relevant spots converted to the geometric model and saved.

It will be appreciated that the present invention encompasses the process described hereinabove for tennis games as well as for other situations in which the background information is fixed and known. The process described hereinabove can be improved in a number of ways, through tracking and through knowledge of the camera parameters, as described hereinbelow.

When information on each camera position, rotation angles and amount of zoom are supplied (either externally or determined by the system), the operations described hereinabove can be shortened since the number of degrees of freedom of the perspective matrix P are reduced.

Specifically, the perspective matrix P includes in it information regarding the position, rotation angles and zoom of the camera utilized. This information can be extracted and the perspective matrix P (or, similarly, the transformation matrix M) can be redefined as a function of each cameras parameters.

Fig. 19 illustrates a camera and its parameters.

Its location is denoted by the vector 171 having coordinates (x,y,z) from the origin O of the X, Y, Z coordinate system 172. The camera rolls, tilts, turns and pans respectively, about camera-based axes U, V and W, as indicated by arrows 173, 174 and 175, respectively. Furthermore, the camera lens can zoom along the V axis, as indicated by arrow 176.

Assuming that the camera does not roll and that the aspect ratio of the camera (the ratio between the width and height of a pixel in the image which the camera produces) defines square pixels, the perspective matrix P can be parameterized as a function of the location (x,y,z) of the camera and of its tilt, turn and zoom. It is assumed that the camera does not change position from frame to frame but just changes its tilt, turn, angles or zoom.

Figs. 20, 21 and 22 represent the method of determining and then utilizing the camera parameters. In Fig. 20, when a new cut is identified in the video flow, the entire process of perspective identification (step 180), as shown in Fig. 10, is performed on the first frame of the new cut. Step 180 produces the $a(i,j)$ elements of the perspective matrix P. The process continues in two directions:

a) The transformation matrix T is determined, starting from step 154 of Fig. 13; and

b) The camera coordinates (x,y,z) are extracted (step 184) from matrix P, as taught in section 3.4 of the book Three-Dimensional Computer Vision: A Geometric Viewpoint, by Oliver Faugeras, MIT Press, 1993. The book is incorporated herein by reference.

Once the camera coordinates (x,y,z) have been extracted, two checks (steps 186 and 188) are performed as follows:

Condition 186: The camera does not roll (rotate) in

direction 174. Roll is present when the element a_{13} is not equal to zero.

Condition 188: The aspect ratio (AR) of the camera defines square pixels. (i.e., $AR=1$)

- 5 If either condition is false, the remainder of the shortened process is aborted.

If both conditions are fulfilled, then, as taught in the book Three-Dimensional Computer Vision: A Geometric Viewpoint, the matrix P can be re-represented
10 (step 190) as the product of the following matrices:

- a) Zoom (f) : the matrix of the projection of the camera focal-plane;
- b) Translation: the matrix of translation from the coordinate system origin to the computed camera
15 position, (x, y, z) ;
- c) Tilt (α): the matrix of the rotation around the U axis through the angle α ; and
- d) Turn (θ): the matrix of the rotation around the W axis through the angle θ .

20 With the values of zoom, tilt, turn and translation, the first camera is fully calibrated (step 192) and its parameters are inserted into a table 194 of the identified cameras (shown in Fig. 21). Other cameras will be identified and registered in table 194 as
25 described hereinbelow.

The shortened calculation process, described with respect to Fig. 22, is then performed on all frames. A frame is examined (step 196) to determine its similarity to the previous frames, using α , θ and f . Similarity is
30 measured via a matching coefficient (i.e., percentage of pixels of interest in the frame successfully mapped to the model using the computed matrix). If a good similarity is obtained, the computed matrix can be used for the insertion process (described with respect to Fig.
35 17). If the matching coefficient is small, it is possible that this frame was filmed by another camera

from table 194.

To find the other camera, the current frame must be reviewed and one line in it must be identified. Furthermore, a point on the identified line, such as an intersection point with another line, must also be identified (step 198). Usually the identified line is the "strongest" line.

In step 200, a match value for each camera listed in table 194 is determined, as follows:

10 The identified line and point are associated with a line and point in the geometric model, and a perspective matrix P for this association is determined which transforms the line and point of the model to the identified line and point. Since each perspective matrix
15 P is a function of the coordinates (x, y, z) of the current camera (which are known) and the tilt α , turn θ and zoom f (which are unknown), the resultant perspective matrix P can be determined through the values of the tilt, turn and zoom which can be computed, assuming that the
20 identified line and point are properly matched to the line and point of the model.

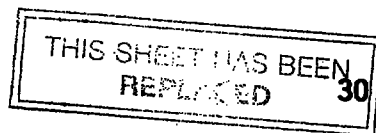
As in the method of Fig. 10, the transformation matrix M is determined from the perspective matrix P and the geometric model is transformed, through the matrix M ,
25 into the plane of the image in the frame. The lines of the model are matched to the lines in the image and a match value produced.

The process of associating a line and point of the model with the identified line and point, producing
30 a perspective matrix P from the known camera coordinates and the association of lines and points, and determining a match value as a result, is repeated for each combination of line and point in the geometric model. If the match values are considerably less than 1, indicating
35 that the match was very poor, the match process with the identified line and point, described hereinabove, is

repeated for another camera whose coordinates (x,y,z) are known.

The largest, computed matching coefficient for each camera is inserted into a column, labeled 202, of table 194 (Fig. 21). In step 204, the camera with the highest value of coefficient 202 is selected, and, if the coefficient is larger than a predefined threshold, its perspective matrix P is used for the image insertion process of Fig. 17. If the highest coefficient in column 202 has a value lower than the threshold, no known camera was used to shoot the current frames. The process of Fig. 10 must be performed followed by the camera identification process of Fig. 20.

It will be appreciated by persons skilled in the art that the present invention is not limited to what has been particularly shown and described hereinabove. Rather the scope of the present invention is defined by the claims which follow:



CLAIMS



1. A system for mixing an image with a video stream of action occurring within a background space, the space having fixed surfaces and being scanned by at least one camera in a non-determinate manner, the system comprising:

a video frame grabber for grabbing a single frame of a video signal at one time;

an image implantation system comprising:

means for receiving a) a flat model of said fixed surfaces and b) an image mask indicating a preselected portion of said flat model onto which said image is to be mixed;

identification means for utilizing said model to determine if and where said preselected portion is shown in said frame; and

implantation means utilizing the output of said identification means for implanting said image into said frame onto said preselected portion of said fixed surfaces, if said preselected portion is shown in said frame.

2. A system according to claim 1 and also comprising a design workstation for previously designing said image and said image mask, said design workstation occasionally communicating with said means for receiving.

3. A system according to claim 1 and wherein said identification means comprises:

means for reviewing said single frame and for extracting features of said fixed surfaces therefrom; and

means for determining a perspective transformation between said model and said extracted features.

This Sheet Has Been Replaced





4. A system according to claim 3 wherein said means
for reviewing and extracting comprises means for creating
a background mask and a foreground mask and extraction
means for utilizing said background mask to extract
5 desired features, wherein said background mask indicates
the locations of features of interest, of background
elements and of foreground elements, wherein said
foreground mask is formed of the foreground data portions
of said background mask, and wherein said extraction
10 means utilize the features of interest portion of said
background mask to extract said desired features.

5. A system according to claim 4 and wherein said
15 means for creating comprises a lookup table for
converting between the multiplicity of colors in said
frame to one of: colors of features of interest, colors
of background elements and a color indicating foreground
elements.

20

6. A system according to claim 5 and wherein said
means for creating comprises means for producing said
lookup table from colors indicated by a user and
correction means for correcting said lookup table when
25 said colors no longer indicate the features of interest
and the background elements.

7. A system according to claim 4 and wherein said
implantation means comprises:

30 means for transforming said image and said image
mask with said perspective transformation; and

means for mixing said transformed image and said
transformed image mask with both said single frame and
with said foreground mask which indicates locations of
35 foreground data not to be covered by said transformed
image.



THIS SHEET HAS BEEN
REPLACED

8. A system according to claim 4 and wherein said implantation means comprises:

means for transforming said image, said image mask and a blending mask with said perspective transformation; and

means for mixing said transformed image, said transformed image mask and said transformed blending mask with both said single frame and said foreground mask which indicates locations of foreground data not to be covered by said transformed image.

9. A system according to claim 4 and wherein said features are lines.

10. A system according to claim 9 and wherein said means for reviewing and extracting comprises a Hough transform.

11. A system according to claim 9 and wherein said means for reviewing and extracting comprises means for determining the angles of segments of said lines by investigating neighborhoods of pixels and by continually narrowing said neighborhoods.

12. A system according to claim 9 and wherein said means for determining comprises means for projecting said extracted features onto an asymptotic function to determine which of said features are perspective versions of parallel lines.

13. A system according to claim 9 wherein said background space is a sports arena and wherein said means for determining also comprises:

a list of rectangles of said model and the locations of their corner points;

means for selecting two vertical and two

This Sheet Has Been Replaced



* C L M S *

THIS SHEET HAS BEEN
REPLACED

PCT/US95/02424

horizontal lines from said extracted features and for determining their intersection points;

means for generating transformation matrices from the corner points of each rectangle of said model to said feature intersection points;

means for transforming said model with each transformation matrix;

means, utilizing said background elements of said background mask, for matching each transformed model with said single frame; and

means for selecting the transformation matrix which matches the features of the frame best.

14. A system according to claim 13 and wherein said means for determining additionally comprises:

means for receiving or extracting the coordinates of a set of cameras;

means for representing a current transformation matrix as a product of coordinate, tilt, turn and zoom matrices and then determining the values for the tilt, turn and zoom; and

means for identifying the camera having the calculated values for tilt, turn and zoom and storing the information; and

means for repeating the steps of receiving, representing and identifying whenever there is a new cut in the video.

15. A method for mixing an image with a video stream of action occurring within a background space, the space having fixed surfaces and being scanned by at least one camera in a non-determinate manner, the method comprising the steps of:

grabbing a single frame of a video signal at one time;

receiving a) a flat model of said fixed surfaces

This Sheet Has Been Replaced





and b) an image mask indicating a preselected portion of said flat model onto which said image is to be mixed;

utilizing said model to determine if and where said preselected portion is shown in said frame; and

5 with the output of said step of utilizing, implanting said image into said frame onto said preselected portion of said fixed surfaces, if said preselected portion is shown in said frame.

10 16. A method according to claim 15 and also comprising the steps of previously designing said image and said image mask and occasionally communicating with said means for receiving.

15 17. A method according to claim 15 and wherein said step of utilizing comprises the steps of:

reviewing said single frame and for extracting features of said fixed surfaces therefrom; and

20 determining a perspective transformation between said model and said extracted features.

18. A method according to claim 17 wherein said step of reviewing and extracting comprises the step of creating a background mask and a foreground mask, wherein
25 said background mask indicates the locations of features of interest, of background elements and of foreground elements, wherein said foreground mask is formed of the foreground data portions of said background mask, and the step of using features of interest portion of said
30 background mask to extract said desired features.

19. A method according to claim 18 and wherein said step of creating includes the step of converting, via a lookup table, between the multiplicity of colors in said
35 frame to one of: colors of features of interest, colors of background elements and a color indicating foreground

This Sheet Has Been Replaced



elements.



20. A method according to claim 19 and wherein said step of creating comprises the step of producing said
5 lookup table from colors indicated by a user and the step of correcting said lookup table when said colors no longer indicate the features of interest and the background elements.
- 10 21. A method according to claim 18 and wherein said step of utilizing comprises the steps of:
transforming said image and said image mask with said perspective transformation; and
mixing said transformed image and said
15 transformed image mask with both said single frame and with said foreground mask which indicates locations of foreground data not to be covered by said transformed image.
- 20 22. A method according to claim 18 and wherein said step of utilizing comprises the steps of:
transforming said image, said image mask and a blending mask with said perspective transformation; and
mixing said transformed image, said transformed
25 image mask and said transformed blending mask with both said single frame and said foreground mask which indicates locations of foreground data not to be covered by said transformed image.
- 30 23. A method according to claim 18 and wherein said features are lines.
24. A method according to claim 23 and wherein said step of reviewing and extracting comprises the step of
35 performing a Hough transform.



25. A method according to claim 23 and wherein said step of reviewing and extracting comprises the step of determining the angles of segments of said lines by investigating neighborhoods of pixels and by continually
5 narrowing said neighborhoods.

26. A method according to claim 23 and wherein said step of determining comprises the step of projecting said extracted features onto an asymptotic function to
10 determine which of said features are perspective versions of parallel lines.

27. A method according to claim 23 wherein said background space is a sports arena and wherein said step
15 of determining also comprises the steps of:

providing a list of rectangles of said model and the locations of their corner points;

selecting two vertical and two horizontal lines from said extracted features and determining their
20 intersection points;

generating transformation matrices from the corner points of each rectangle of said model to said feature intersection points;

transforming said model with each transformation
25 matrix;

utilizing said background elements of said background mask to match each transformed model with said single frame; and

selecting the transformation matrix which matches
30 the features of the frame best.

28. A method according to claim 27 and wherein said step of determining additionally comprises the steps of:

receiving or extracting the coordinates of a set
35 of cameras;

representing a current transformation matrix as

This Sheet Has Been Replaced



* C L M S *

a product of coordinate, tilt, turn and zoom matrices and then determining the values for the tilt, turn and zoom; and

5 identifying the camera having the calculated values for tilt, turn and zoom and storing the information; and

repeating the steps of receiving, representing and identifying whenever there is a new cut in the video.

10

This Sheet Has Been Replaced



* C L M S *



AMENDED CLAIMS

[received by the International Bureau on 24 August 1995 (24.08.95);
original claims 1-28 replaced by new claims 29-36 (2 pages)]

- 5 29. A method for implanting an image into a selected
one at a time of a plurality of video frames
representing a stream of action occurring within a
background space, the space having fixed planar
surfaces and being scanned by at least one video
10 camera, the method comprising the steps of:
generating a model, independent of said plurality
of video frames, of a selected one of said fixed
surfaces, said model comprising a representation of
geometrical features characterizing said surface; and
15 utilizing said model for implanting said image
into said frames, said step of utilizing comprising the
step of perspectively distorting said model.
30. A method according to claim 29 wherein said
20 geometrical features comprises at least one of the
group consisting of lines and arcs.
31. A method according to claim 30, further comprising
the step of providing an indication of the planar
25 relationship between individual ones of said lines and
arcs.
32. A method according to claim 29 wherein said
representation is a planar vectorial representation.
- 30 33. Apparatus for implanting an image into a selected
one at a time of a plurality of video frames
representing a stream of action occurring within a
background space, the space having fixed planar
35 surfaces and being scanned by at least one video
camera, comprising:
means for generating a model, independent of said
plurality of video frames, of a selected one of said

96 1 12 81 - 8



fixed surfaces, said model comprising a representation of geometrical features characterizing said surface; and

5 means for utilizing said model for implanting said image into said frames, said means for utilizing comprising means for perspectively distorting said model.

10 34. Apparatus according to claim 33 wherein said geometrical features comprises at least one of the group consisting of lines and arcs.

15 35. Apparatus according to claim 34, further comprising means for providing an indication of the planar relationship between individual ones of said lines and arcs.

20 36. Apparatus according to claim 33 wherein said representation is a planar vectorial representation.