Visualization System for Player's Sequential Forms at Sports Broadcasting

Sang Gil Lee*, Nobuyuki Yagi**

*Technical Research Institute, KBS
18, Yoido-dong, Youngdungpo-ku, Seoul, 150-790, Korea
tel : +82-2-781-5971, fax : +82-2-781-5999, email : sglee@tri.kbs.co.kr
**Science and Technical Research Laboratories, NHK
1-10-11, Kinuta, Setagaya-ku, Tokyo, 157, Japan
tel : +81-3-5494-2301, fax : +81-3-5494-2309, email : yagi@strl.nhk.or.jp

ABSTRACT : *We developed a visualization system for sports broadcasting that informs TV viewers of a sports player's sequential performing forms and locus in detail when gymnastic game such as 'vault' is broadcasted. It records the corresponding frames while a player is performing, extracts only the player's performing forms from background scenes by a proposed algorithm, superposes them sequentially on a background scene, and broadcasts them with slow speed. This system named as 'Multi-Motion' has ever been used in NHK live broadcasting for 1995 World Gymnastics Championships held in Japan.*

1. INTRODUCTION

In case of conventional broadcasting for sports game such as gymnastics, a player's performance are recorded from a camera to a VTR, and then the scenes recorded are replayed with slow speed for TV viewers to look at the player's sequential performing forms in detail. However, most of TV viewers cannot perceive them well because only one form at a time is displayed sequentially. They often want to observe a player's all performing forms at a time on a scene from initial to final position.

We developed a visualization system that can accomplish such a function in a 'vault' game. While a player is performing after jumping as shown in Figure 1(a), the performing forms are taken sequentially by the camera, and recorded in a HDS(Hard Disk System) with real speed. The desired frames are transferred to a workstation with high speed to be processed. Only the player's performing forms are extracted from the corresponding frames by a proposed algorithm, and they are sequentially superposed on a background scene as shown in Figure 1(b). The resultant frames are broadcasted.

In order to realize such a system, we should extract a moving object from the background scene including the moving objects because the audiences behind the performing player are moving. Besides, we should consider several conditions to operate the system for live broadcasting, that is, processing time, simple and easy operation, and compactness.

This system named as 'Multi-Motion' was used in NHK live broadcasting for the 1995 World Gymnastics Championships held in Japan in October 1995.

We will explain system environment and proposed algorithm in Section 2 and Section 3, respectively. The implementation will be depicted in Section 4, and conclusions will be remarked finally in Section 5.

2. SYSTEM ENVIRONMENT

The proposed system consists of a camera, a real-time HDS, and a workstation as shown in Figure 2. The camera is located in the gymnastic stadium, and the other outside the stadium. The production facilities might be mobile production unit or a temporary building where program is producted. The desired results are generated via the following stages.

- **Recording Stage** : A camera is located in front of the place where the vault game is done. An operator gives 'Record' command to

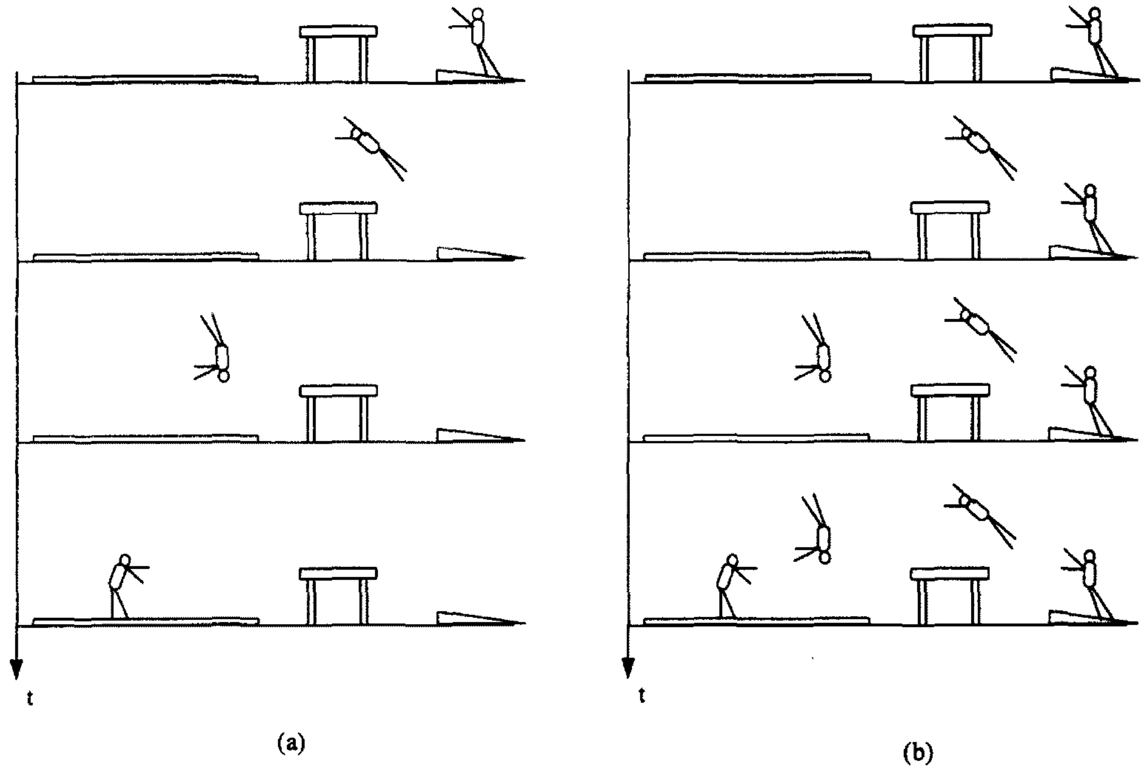

Figure 1. The example that shows TV viewers the player's sequential forms after performance. (a) The player is shown one form at a time by the conventional method. (b) The player's sequential forms are shown simultaneously on a scene.

the system when the player is just going to pass through the jumping plate. The system starts to record all frames from the camera into the HDS until the player finishes performance.

• **Execution Stage** : We search a reference frame recorded in the HDS. The reference frame means the frame in which no player exists just before the player starts jumping. The frames are transferred to the workstation starting from the selected reference frame. The transferring speed between the HDS and the workstation should be faster enough for the system to perform the live broadcasting. The proposed algorithm handles image files on the workstation to extract only the player's performing forms from the background. And the player's all forms are superposed one at a time along the sequential frame. The proposed algorithm will be depicted in the next Section.

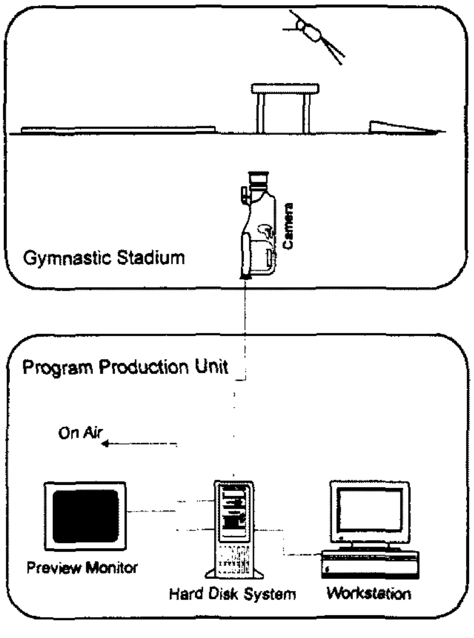

Figure 2. System Environment.

• **Display Stage** : The output frames obtained at execution stage are transferred into the HDS again, and broadcasted. Resultantly TV viewers perceive the sequential flow of the performing forms in detail.

3. PROPOSED ALGORITHM

Among the frames transferred from the HDS to the workstation, the first frame is regarded as the reference frame in which a performing player does not exist. We extract the moving objects by getting difference between the reference frame and the current frame. If we fix the position of the reference frame to the first frame, the distance between the reference frame and the current frame becomes larger as we process the algorithm along the frames. It results in the higher occurrences of undesired moving objects. This problem can be solved by limiting the processing area of the current frame to the square area surrounding the moving player, and by selecting the reference frame from one of the previous frames as close as possible to the current frame. Changing the processing area and reference frame, we continue to process the algorithm till the last frame along the following steps.

3.1. Prediction of processing area

The player's jumping position is already known, and the locus of the player can be predicted. Therefore the processing area can be limited to the small area that surrounds the performing player at the starting frame. The position and size of the processing area can be predicted according to the player's movement from the starting frame by location, moving direction, and moving speed of the desired moving object at the previous frame processed.

Let $B_{n-1}(x,y)$ be a binary key signal indicating the desired moving object extracted and PA_{n-1} be the processing area at $(n-1)th$ frame, respectively. PA_{n-1} is the smallest area which includes $B_{n-1}(x,y)$. And we set the center position of $B_{n-1}(x,y)$ as (cx_{n-1}, cy_{n-1}), the horizontal and vertical size of PA_{n-1} as (wx_{n-1}, wy_{n-1}), and the displacement of the desired moving object from $(n-2)th$ to $(n-1)th$ frame as (dx_{n-1}, dy_{n-1}). In this case, the center position and the size of the processing area at the current frame become $(cx_{n-1}+dx_{n-1}, cy_{n-1}+dy_{n-1})$, and $(wx_{n-1}+\alpha x_n, wy_{n-1}+\alpha y_n)$, respectively. Here αx_n and αy_n are horizontal and vertical tolerance, which are positive and are set large enough to include $B_n(x,y)$ to be derived. The limitation of the processing area can also save processing time greatly.

3.2. Selection of reference frame

A reference frame is defined as the frame which has no player inside the processing area. Using the predicted processing area, we select the reference frame which is used to obtain difference with the current frame. Assuming that the moving objects do not appear except a player, the shape of binary key signal obtained from the difference between the current frame in Figure 3(c) and the reference frame depends on the selection of the reference frame. If the reference frame does not include the performing player inside the processing area as shown in Figure 3(a), the shape of binary signal is obtained correctly as shown in Figure 3(d). Otherwise, erroneous binary key signal is obtained as shown in Figure 3(e). Therefore the distance between the reference frame and the current frame should be large enough for reference frame not to include the performing player inside the processing area.

Practically, the background at a sports scene includes many moving objects except the player because audiences are moving behind the performing player. So undesired moving objects might be detected as the moving objects. The closer the distance between the reference frame and the current frame is, the smaller the occurrences of the undesired moving objects are. As the occurrences of undesired moving objects become smaller, the distortion of desired moving object is reduced. Therefore we select the frame as a reference frame, which is closest to the current frame among the previous frames where no player appears inside the predicted processing area.

3.3. Extraction of moving object

Assuming that the pth previous frame from the current frame with its index being n is selected as the reference frame, we will derive the moving objects, $B_{n,n-p}(x,y)$, between the nth frame and the $(n-p)th$ frame. Firstly, each pixel is averaged spatially in order to be smoothed by spatial averaging mask with the size being 3x3. It reduces the occurrences of the non-zero difference pixels larger than arbitrary threshold to be detected as the undesired moving object. Let $F_n(x,y)$ and $F_{n-p}(x,y)$ be the outputs of spatial masking for the pixels of the current frame and the reference frame inside the processing area, respectively. The difference between the current frame and the reference frame is derived from *Eq.* (1), and is converted to a binary signal by *Eq.* (2).

$$D_{n,n-p}(x,y) = |F_n(x,y) - F_{n-p}(x,y)| \quad (1)$$

$$B_{n,n-p}(x,y) = \begin{cases} 1 & \text{if } D_{n,n-p}(x,y) \geq T_m \\ 0 & otherwise \end{cases} \quad (2)$$

where '1' and '0' indicate the moving and still pixel, respectively. These binary objects

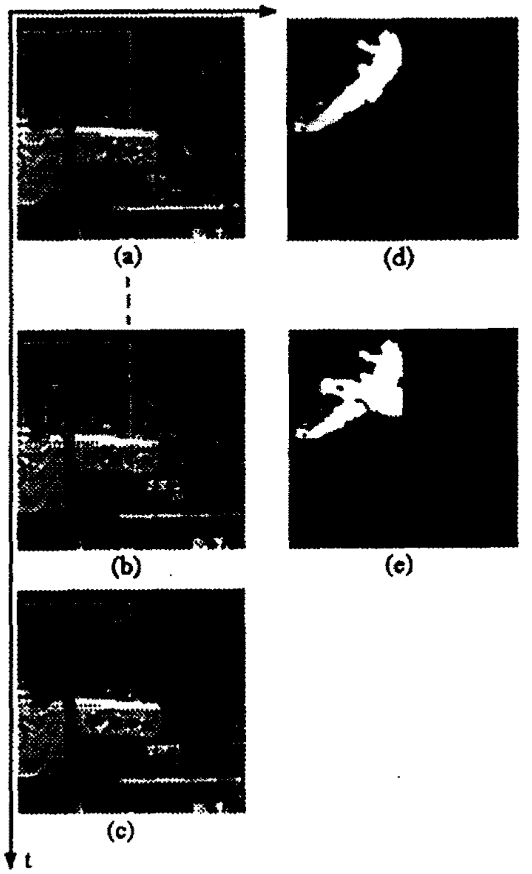

Figure 3. Binary images obtained according to the selection of reference frame.
(a) The reference frame selected correctly for the player not to overlap with the player at current frame inside processing area. (b) The reference frame selected incorrectly for the player to overlap with the player at current frame inside processing area. (c) Current frame to be processed inside the processing area. (d) Correct binary image obtained by correct selection of reference frame. (e) Erroneous binary image obtained by incorrect selection of reference frame.

will be classified into the largest binary object and another objects through labeling and sorting. Different number is allocated to each independent moving object orderly. And then the largest moving objects is selected among all labeled moving objects. The largest binary object includes the desired moving object.

3.4. Updating processing area, reference frame, and moving object

The largest moving object obtained still includes the undesired moving objects. We reduce the processing area, and select a new reference frame for derivation of more accurate moving object.

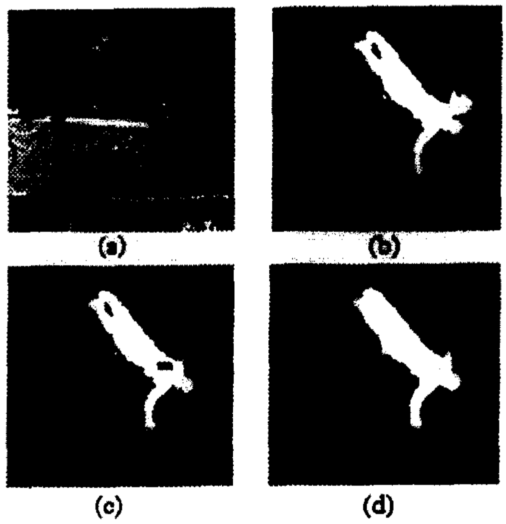

Figure 4. Status at each step during process.
(a) original image. (b) The moving objects are extracted. (c) The undesired moving object is removed. (d) The blocky effect is removed.

The processing area can be updated by the moving object obtained before. That is, the size of the processing area can be reduced to the smallest area which includes the moving object. The horizontal and vertical size of the updated processing area will be smaller than $wx_{n-1}+\alpha x_n$ and $wy_{n-1}+\alpha y_n$, respectively. Using the updated processing area, the reference frame with its index being p can be changed to the frame with its index being q under condition with $p \leq q \leq n$. Finally more exact moving object is extracted using new reference frame inside the updated processing area and reference frame. However, the obtained binary moving object still includes the binary data of the undesired moving objects as shown in Figure 4(b).

3.5. Removal of undesired moving object

The undesired moving object will be removed by motion compensation and adaptive threshold depending on frequency characteristics. The player exists not at the reference frame but the current frame inside the processing area by the given condition. On the other hand, the undesired moving objects might almost exist at both of the reference frame and the current frame.

Assuming that the blocks are translated rigidly from the reference frame to the current frame, the pixels corresponding to the undesired moving object at the current frame can be found at the reference frame by motion estimation. However the pixels of the desired moving objects cannot be found at the reference frame because its motion vector, (v_x,v_y), is false.

From the above assumption, the sum of the difference between the block of the current frame and the motion compensated block of the reference frame will not be zero at the desired moving object, but zero at the undesired moving object theoretically. Removing the blocks of which the sums of the difference are zero, the undesired moving objects can be removed from the desired moving object. Practically, the sum of the difference of the undesired moving block will be less than or equal to the arbitrary threshold, T_{mc}, and the sum of the difference of the desired moving block will be more than T_{mc}.

Although the motion vectors for the undesired moving object are true, there might appear the blocks at which the sums of the difference between the blocks of the current frame and the motion compensated blocks of the reference frame are more than T_{mc}. So those blocks at undesired moving part are not removed occasionally.

The sum of the difference at the block which has a true motion vector is proportional to the high frequency component at the corresponding block of the current frame. That is, the more complex the block is, the larger the sum of difference is. We should adapt the higher threshold value to such undesired moving blocks to remove it because the block including audiences has high frequency components statistically. Therefore we adapt the threshold, $T_{mc}(w)$, depending on the characteristics of frequency.

Eq. (3) measures the quantity of horizontal or vertical high frequency components for an arbitrary block by high pass filter, $H_{hp}(w)$, where coefficient is [-1/2, 1, -1/2] horizontally or vertically. $T_{mc}(w)$ is obtained by *Eq.* (4).

$$F_h(w) = H_{hp}(w) \cdot F(w) \tag{3}$$

$$T_{mc}(w) = \alpha \cdot F_h(w) + \beta \tag{4}$$

where α and β are constants. *Eq.* (5) shows the sum of the difference at the block of the current frame and the motion compensated block of reference frame.

$$E_{n,n-q}(x,y) = \sum_{k=0}^{N}\sum_{l=0}^{N} |F_n(x-k, y-l) - F_{n-q}(x-k-v_x, y-l-v_y)| \tag{5}$$

where N is block size. The removal of the undesired moving block from the binary moving object obtained is performed with block unit by *Eq.* (6) while the desired moving blocks are preserved.

$$B_{n,n-1}(x,y) = \begin{cases} 0 & \text{if } E_{n,n-q}(x,y) \le T_{mc}(w) \\ 1 & otherwise \end{cases} \tag{6}$$

The process using *Eq.* (3) and *Eq.* (4) is performed horizontally at first, and vertically.

3.6. Removal of blocky effect

The binary key image signal, $B_{n,n-1}(x,y)$, obtained by *Eq.* (6) has the blocky effect because removal of the undesired moving blocks are performed with block unit. Therefore the shape of the extracted player's forms might not be natural as shown in Figure 3(c). For pixel with its value being zero inside the processing area, we check the number of non-zero pixels within $(N+1)$x$(N+1)$ pixels surrounding the corresponding pixel. If the number is larger than or equal to $2N$, it will be non-zero. By this process, the blocky effect disappears. Besides, it is often necessary to fill holes which might occur during removal of the undesired moving object. Figure 4(d) shows the result after the removal of the blocky effect and the hole filling.

4. IMPLEMENTATION

The system consists of a camera, a real-time hard disk system, and a workstation as explained in Section 2.

In order to save processing time, all frames whose size is 720x480 are subsampled horizontally and vertically to be reduced to half size at each direction. We used color components as well as luminance component in order to extract moving objects from background effectively. T_{mc} used in *Eq.* (2) depends on the light condition in the gymnastic stadium.

We set parameters for motion estimation as follows.

- measurement block size : 4V x 4H
- range of searching area : 32V x 16H

The threshold used to remove the undesired moving object from the moving object is changed according to the frequency characteristics of luminance signal in the corresponding measurement block of the current frame. We optimized the coefficients α and β. Because the binary key signals are obtained from the subsampled frame, the extracted binary key signals are interpolated. The interpolated binary key signal has the blocky effect because block motion estimation is used. By the removal method of blocky effect proposed, the soft binary key signals are obtained. Figure 5 shows the results of a player's partial sequential forms extracted from the original image sequences. These sequential player's forms are superposed on a background scene, and shown to TV viewers as shown in Figure 6.

5. CONCLUSIONS

We developed the visualization system to show TV viewers a sports player's performing forms in detail. This system named as 'Multi-Motion' was used in NHK live broadcasting for the 1995 World Gymnastics Championships held in Sabae, Japan from Oct. 1 to Oct. 10, 1995. It can be applicable to another kinds of live sports broadcasting with further studies on the updated extraction methods which are robust to any background characteristics.

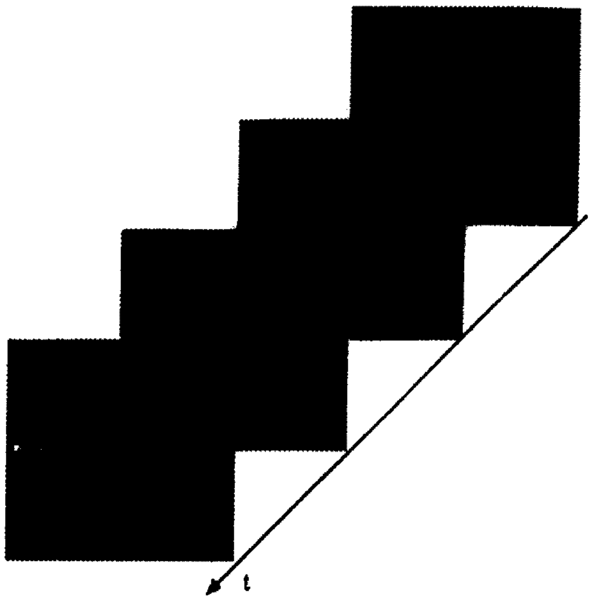

Figure 5. Results of a player's partial sequential forms extracted.

Figure 6. An image scene visualized by superposition of a player's sequential forms extracted from image sequences.

REFERENCES

1. S.D. Blostein and T. S. Huang, "Detecting small, moving objects in image sequences using sequential hypothesis testing," *IEEE Trans. on Signal Processing*, vol. 39, no. 7, pp. 1611-1629, Jul. 1991.
2. M. Bichsel, "Segmenting simply connected moving objects in a static scene," *IEEE Trans. on Pattern Anal. Machine Intell.*, vol. 16, no. 11, pp. 1138-1142, Nov. 1994.
3. N. Mine, et al., "Detection of Change Region by Integrating Subtracted Image and Edge Boundary Image," *IEICE D-II*, vol. J77-D-II, no. 3, pp. 631-634, 1994.
4. A. Broggi, "Parallel and local feature extraction : A real-time approach to road boundary detection," *IEEE Trans. on Image Processing*, vol. 4, no. 2, pp. 217-223, Feb. 1995.