# An Intelligent Robotic Camera System

**By Noritada Yamanaka, Yoshikazu Yamamura, and Kazuyuki Mitsuzuka**

*An intelligent robotic camera system for broadcast use has been developed that automatically follows a moving object in the TV picture. The video output signals from this camera are processed to generate movement data required by the pan-and-tilt head and the lens servo system. The system features real-time following operation, and desirable camera shots are available at any time.*

Robotic cameras mounted on remote-controlled pan-and-tilt heads are currently in use at NHK studios for news and educational programs. Usually a single (but sometimes more than one) robotic camera is used among the other studio cameras. Such cameras are becoming indispensable in improving the efficiency of program production.

In the robotic camera system, the horizontal and vertical swing angles of the camera and focus positions for zooming are memorized. Accurate recall for later operation is obtained by manipulating the relevant control buttons. The system, however, cannot follow unexpected variations in the picture composition due to slippage and other such shifts in the position of the object or camera. This can sometimes produce unnatural effects in the picked-up image, such as where the head of a player moves out of the frame, even when he moves only slightly.

An intelligent robotic camera has now been developed to eliminate this drawback. The camera detects slippage of the object in the frame and corrects the camera angle automatically. The system can make the camera change its angle automatically to follow the object. It achieves this by adding an image-processing device to the existing system without altering the conventional pan-and-tilt head, thus providing an economical and easy way to operate the robotic camera.

## Conditions for Development

This system had to meet the stringent requirements of broadcasting stations. The following conditions were considered at the development stage.

### *Correct Reproduction of Image Composition*

The object to be picked up may appear not only in the center of the frame, but anywhere within it, according to the purpose of the shot. The position of the object to be referenced later must therefore be recordable at any place in the frame.

It is important that the composition of the frame be reproduced accurately when required, as the aim of the system is automatic compensation for image composition. The position of the object must be detected correctly. It was decided that the position of the object should be detected by the camera itself and defined by processing of the image picked up, and not by external devices such as infrared rays, sensors, and the like.

### *Smooth Movement*

To realize the smooth movement of the object in a frame, it is necessary to process the image quickly and to control it as precisely as possible. It was therefore essential to speed up image processing, the longest part of the process.

### *Securing a Stable Image*

A good cameraman moves his camera as little as possible, even if the object moves suddenly, because he can predict the next movement from experience. Even if a robotic camera can follow the object's movement as described, image instability will result if the camera moves too often. An algorithm has been developed to prevent the camera from moving too much.

### *Easy Operation*

Even with automatic image compensation, easy operation is required in broadcasting stations. The policy was thus adopted of not requiring special operational procedures for the intelligent system. Instead, the operational procedure was to be the same as that of the conventional robotic camera, utilizing the existing operation panel.

## Operational Principle of Intelligent Robotic Camera System

The operational principle of this system is shown in Fig. 1. Specific color components of the object in the frame are extracted from the output image of the camera, and the color of the subject's face is selected as the key. An image is processed to determine the center of gravity of the face; the center of a waist shot is determined from the reference image. If the person moves slightly, the present output image will become an unnatural image and at the same time a new center of the face color will be detected. Then a new center of the present output image will be compared with that of the reference image. By such a comparison, the calculation of vector to be compensated is obtained. Based on this data and by keeping the image centered, the pan-and-tilt head follows the subject in a constant position.

### *System Composition*

The composition of the hardware system of the intelligent robotic camera is shown in Fig. 2. It is composed fundamentally of a camera device, a key signal processor, an image signal processor, a control device, and a remote-controlled pan-and-tilt head. The camera, remote-controlled pan-and-tilt head, pan-and-tilt head controller, and pan-and-tilt head control panel are the conventional units. The system function is realized through the addition of a key signal processor, image signal processor, and controlling device.

### *Key Signal Processor*

The key signal processor serves as a kind of filter when the image from the camera is input to the image signal processor board. In this system, the signal corresponding to the object is dis-

Presented at the SMPTE Advanced Television and Electronic Imaging Conference in Chicago (paper no. 28-26) on February 5,1994. Noritada Yamanaka, Yoshikazu Yamamura, and Kazuyuki Mitsuzuka are with NHK (Japan Broadcasting Corp.), Tokyo, 150-01, Japan. An unedited version of this article appears in *Proceedings of the SMPTE Advanced Television and Electronic Imaging Conference*, SMPTE, 1994. 

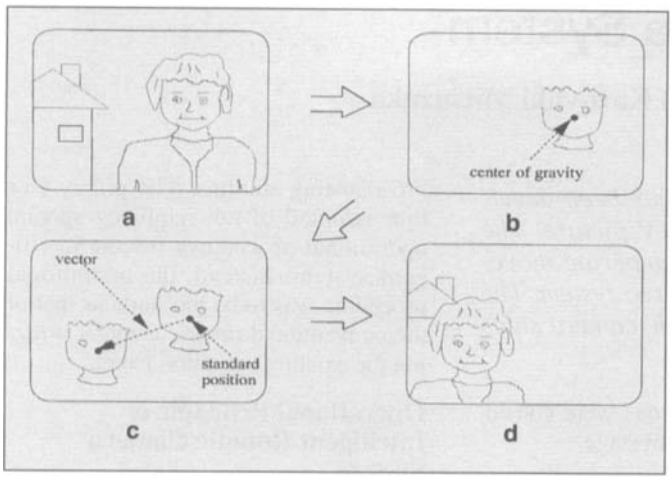


*Figure 1. Principle of intelligent robotic camera system: (a) output image of camera (reference image); (b) extract skin color portion center of gravity; (c) output image of camera when the object moves; (d) compensation value is calculated, comparing its gravity with that of the reference image.*

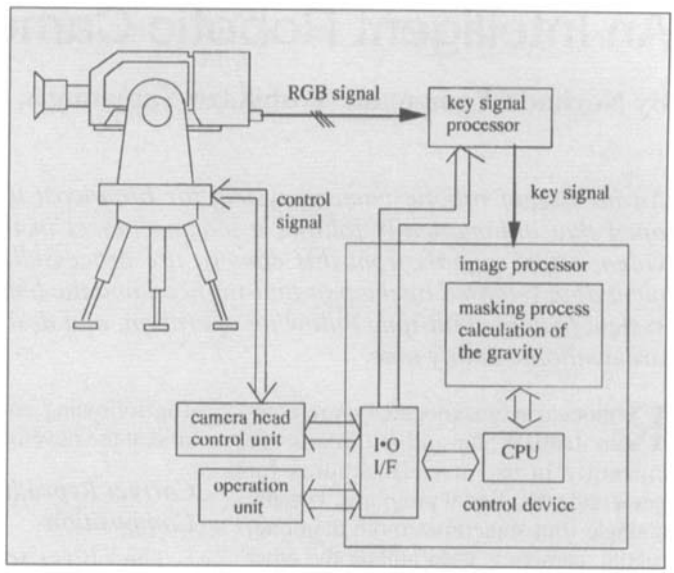


*Figure 2. Hardware composition of intelligent camera system.*

tinguished and extracted from the original image signal by the difference of chromatic components. The specific chromatic component is extracted from the RGB signal output of the camera as a key signal and input to the image processor board.

It is generally thought that images of people are the most commonly used key objects. A human being is usually characterized by skin color, and it is customary to extract a key signal of that color by adjusting a chroma-key device. In cases where an object other than a human being (for example, a flip) is used as a key signal, up to eight kinds of color for keying can be preset. These can be switched freely by remote control.

### *Image Signal Processor*

The image signal processor stores the key signal in memory and determines the center of gravity of the object. Normally systems would store the image signal of a single frame in the memory of the image processor. The image signal of only the first field is stored in this system, so as to shorten the processing period. This is possible because the NTSC system uses an interlace system. Accordingly, the residual vertical blanking period and second field period can be used for calculating the center of gravity and controlling the remote-controlled pan-and-tilt head. The entire process can be finished during the period of a single frame.

A multifunctional image signal processor is used. Measures for noise reduction and labeling are possible, but these would lengthen the processing time by several frames if employed. They would improve the reliability of extraction of the object, but the requirement of real-time processing is more important in this system.

The process is thus simplified by limiting it to two types: determination of the center of gravity and the masking process. The calculation of the center of gravity is less affected by the admixture of a fine noise component into the input signal. Stable operation can thus be expected if the noise is caused by fluctuations in lighting conditions or slight color variation of the object.

### *Controlling Device*

The values of the movement in the X and Y directions, and the resultant total, are fed back to the board from the image signal processor. This is used to obtain the present values of both coordinates of the gravity.

Comparing these with the reference values of gravity for both coordinates, speed-control signals are produced in the pan-and-tilt directions so that the values of the present center of gravity are equal to the reference values of the center of gravity as preset. To achieve smooth control, variations in the values of the gravity position are used as the basis of the speed data for controlling the pan-and-tilt head. These data are combined with the control signals between the operation panel and pan-and-tilt head controller and regulate the pan-and-tilt head directly.

When a shot memory is to be altered, a preset channel is selected that corresponds to this requirement. The color of the key signal is switched to that of the newly selected object.

### ***Operation***

Figure 3 shows the composition of the control data. There are two operational modes; one is a Teach mode for registering the reference position of the object and the shot memories for the pan-and-tilt head, and the other is a Shot Operation mode for reproducing one of the registered shot memories for the pan-and-tilt head to follow automatically.

### ***Registration of Reference Position (Teach)***

Necessary shots are registered during rehearsals, and the shot memories are switched according to the needs of the cut in actual program production. In the intelligent robotic camera, the coordinates of the gravity of the object are automatically calculated and registered as a reference position for the shot at the same time that the shot memory is registered. The key signal processor has channels that correspond to eight different chromatic components, and one of these channels is switched to correspond to the color of the object. The channel number used with the shot memory is also registered automatically to correspond with each shot memory, together with the reference position of the object.

### ***Shot Operation***

When a shot number is selected on the operation panel, the control device

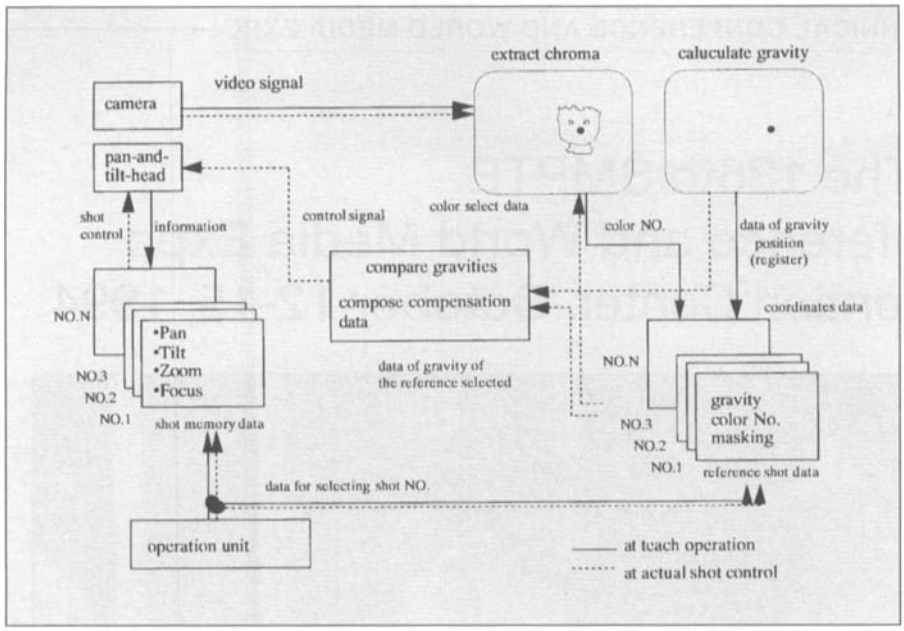


*Figure 3. Composition of control data.*

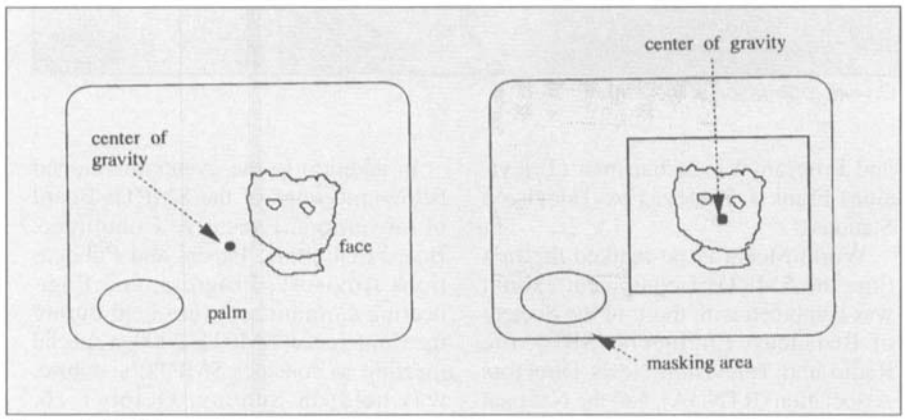


*Figure 4. Masking process.*

sends a shot-selection command to the pan-and-tilt head controller. The pan-and-tilt head recalls the selected shot memory and moves the camera towards the object accordingly. When the pan-and-tilt head begins to move the camera, the control device loads the program and the data for the reference position into the image signal processor. At the same time, it transmits a signal for switching a channel to the key signal processor. When the movement of the camera is finished, the automatic center of gravity detection compensates the frame composition automatically.

### *Masking Process*

When the object of the shot is a human being, the skin color of the person's face is extracted as the key signal to detect the center of gravity (Fig. 4). If there is any other portion of the same skin color, such as a hand, in addition to the face, the data for the center of gravity will be affected by it. To remove this unwanted effect from the key signal component, the masking process conceals the unwanted portion that corresponds to the shot when the image signal is processed.

### *Two-Shot Mode*

When the object to be picked up consists of two people, they may move not only horizontally or vertically, but also away from or towards each other. If the distance increases, one or both of their images may go out of the frame. The distance between the two objects cannot be detected by the process of the center of gravity, and it is impossible to compensate the composition by zooming. Two image signal processing boards are used, one of which is assigned to the left object and the other to the right. The distance between the two objects is measured by dividing the frame into two parts. Two images can be kept from going out of the frame by zooming out when the distance increases, and by zooming in when it decreases.

### *Insensible Area*

If the pan-and-tilt head follows every slight variation of the object's movement, the background of the image will fluctuate constantly and make the image unstable. To make it stable, a rather broad area was designated to be insensible to movement. However, if this area were constantly broad, the sensitivity of follow-up would have to be low. This problem was resolved by setting different insensible areas, depending on the state of the pan-and-tilt head. A broad insensible area is designated when the pan-and-tilt head is not moving, and a narrower area when it is. If the insensible area is too narrow while the pan-and-tilt head is moving, there may be hunting when it stops. The switching of the breadths is made automatically, depending only on the distance between the object and its reference position. The insensible area becomes broader as the object, drawing near to the reference position, enters the area. This effectively prevented hunting.

### *Noise Elimination*

The input image signal is processed after conversion to a binary value in the image signal processor. The fine-noise component included in the input key signal can thus be eliminated there. Even if the noise component remains to some extent, its effect on accuracy will be comparatively small because the process is used only to define the location of the gravity.

## Application and Potential

The present intelligent robotic camera system can be used on similar pan-and-tilt heads from other manufacturers. The system is applicable to any robotic camera currently in use, and it is expected to be used widely in television talk shows, news programs, and so on. Besides these applications, it is also possible to use it for various unmanned operations, such as the photographing of nocturnal creatures, in combination with infrared or other highly sensitive cameras. The system cannot surpass the ability of a human cameraman at present, but the time may come when some images can only be picked up by intelligent robotic cameras.