Live-image Handling for Computer Assisted Presentation

Masashi TODA Takahito YAMAGUCHI Kazuyoshi KANAYAMA
Toshio KAWASHIMA Yoshinao AOKI

Graduate School of Electronics and Information Engineering, Hokkaido University
N13 W8, Kita-ku, Sapporo 060-0813 Japan
E-mail: toda@media.eng.hokudai.ac.jp

Abstract

In usual CCD camera-based presentation systems, a zoom of a camera and a position of a document are manually adjusted for focus by the user. In an intelligent presentation system, however, the zoom parameters of the camera and the area to be focused are automatically determined. Furthermore, a document should be captured in inhomogeneous resolution, because a document image consists of some blocks that are printed in different-sized fonts. In this study, we propose a technique for camera-based inhomogeneous image capturing in order to develop an intelligent presentation system. Zoom parameters and the area to be focused are determined by analysis of the distribution and changes in characteristic vectors, and the images captured are merged using an image-mosaicing technique. An experiment was conducted on live images.

1 Introduction

CCD camera-based presentation systems are widely used (Figure 1). In such systems, a zoom of a camera and a position of a document are manually adjusted for focus by a user (Figure 2). However, automatic adjustment is desirable. So we consider to use an active camera in which we can control the zoom and glance direction (Figure 3) [1].

Document images consist of blocks that are printed in different-sized fonts such as titles, figures, bodies, and footnotes. The image resolution that is needed for presentation depends on the sizes of fonts used in these blocks. In such a case, the document should be captured in inhomogeneous resolution (Figure 4).

There are three technical problems that must be solved in order to realize such an image capturing system:

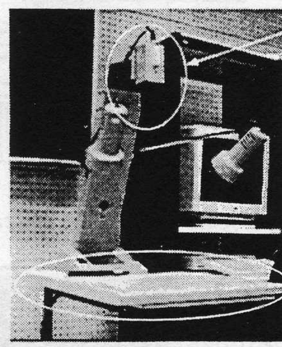


Figure 1: **CCD camera-based presentation system.**

1. Identification of areas that require more detail, that is,
   - extraction of pixels that lack resolution
   - generation of block regions based on the distribution of such pixels
2. Determination of an adequate zoom parameter
3. Overlapping images by voluntary size

In this paper, we propose a solution to these problems.

This type of system, the inhomogeneous image capturing system, is one sub-system of the intelligent presentation system which we are currently developing[1]. The intelligent presentation system can handle bodies, figures, and so on in a paper document (Figure 5).

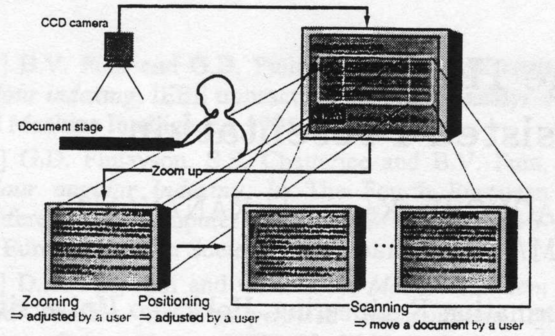


Figure 2: **Former system and problems.**

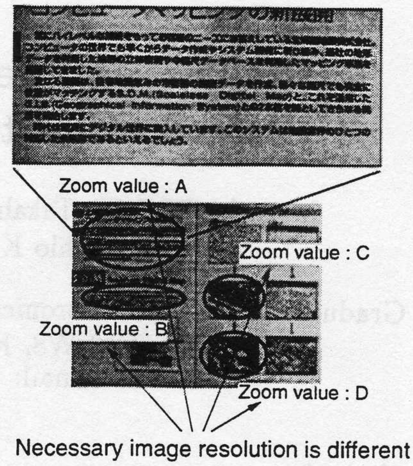


Figure 4: **Inhomogeneous (adaptive) image capturing.**

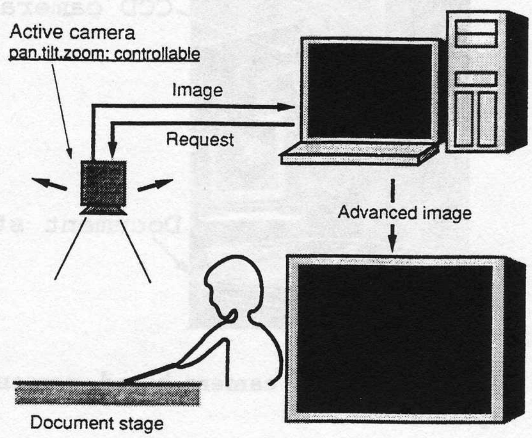


Figure 3: **Conceptual figure of our proposed system.**

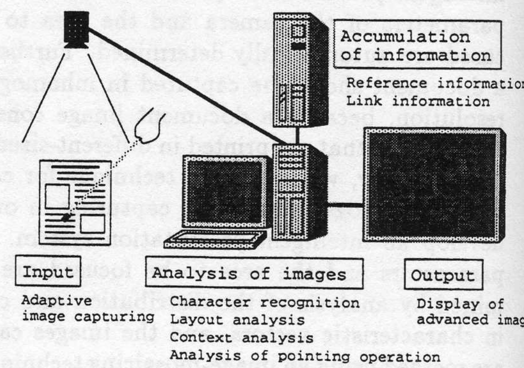


Figure 5: **Intelligent presentation system.**

2 Inhomogeneous image capturing system

2.1 Outline of inhomogeneous image capturing

Figure 6 shows the flow of inhomogeneous image capturing.

First, pixels that require more resolution are extracted. We call such a pixel a zoom-lacking pixel (ZLP). Next, block regions based on the distribution of ZLPs are generated. Then, the user chooses one block for presentation, and an appropriate zoom parameter is determined for the selected block. A series of images is captured by scanning and is then composed.

2.2 Extration of zoom-lacking pixels

A block in Section 1 consists of structual elements such as letters. These structual elements are further divided into pixels.

2.2.1 Property of zoom-lacking pixels

If the zoom value is too low, the structual element is fused with other structual elements or with the background, and the characteristic vectors of pixels are different from the original ones (Figure 7, **state (1)**). If the zoom value is sufficiently high, the structual element is clearly separated from other structual elements, and the characteristic vectors of pixels become the original ones (Figure 7, state (2)).

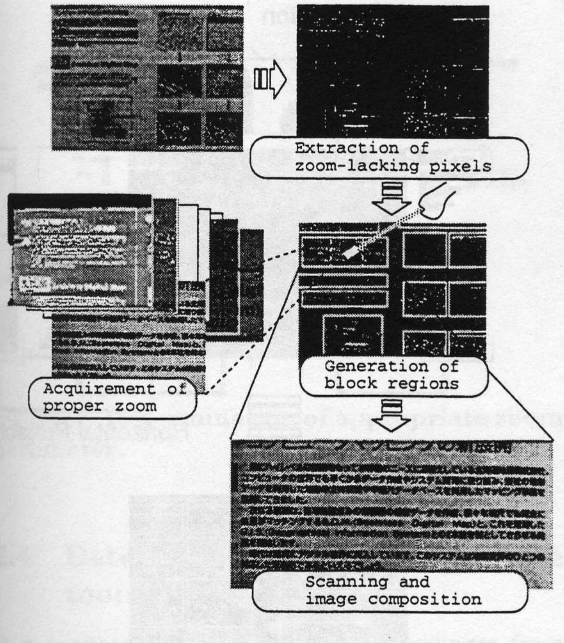


Figure 6: **Flow of inhomogeneous image capturing.**

Property 1 : We classify pixels into clusters by the probability density of characteristic vectors. In state (1), a pixel in the right-middle structual element that lacks resolution is perceived as an outlier, not as a principal cluster. In state (2), the pixel in this structual element is expected to be perceived as a principal cluster. So we notice pixels that are perceived as an outlier.

Property 2 : In state (1), the characteristic vector of the pixel in the right-middle structual element changes greatly with an increase in the zoom parameter. In state (2), however, there is little change in the characteristic vector. So we notice this type of pixel.

We attempted to extract ZLPs using these two properties. Figure 8 shows the flow of extraction of ZLPs. The following is a brief description.

2.2.2 Analysis of distribution of characteristic vectors

We classify pixels into clusters using the improved method[2] of the Maximum Likelihood Robust Clustering (MARC) algorithm[3][4]. By the way, the border of the structual element is always perceived as an outlier and should be eliminated. We call pixels satisfying this condition isolating outlier pixels.

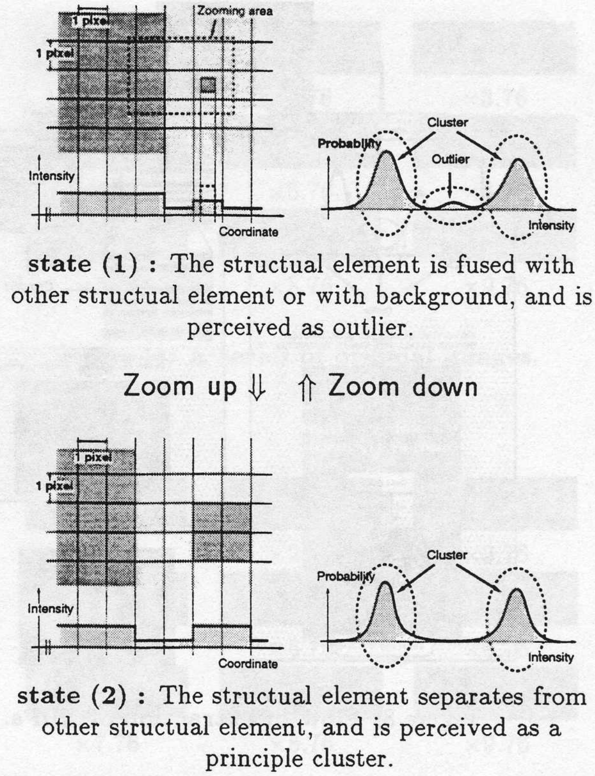


Figure 7: **Relationship between the zoom parameter and the state of the structual element.**

Isolating outlier pixels are considered to be possible ZLPs.

As a weight concerning a possible ZLP through this analysis, I_{uv} is defined.

$$I_{uv} = \begin{cases} 1, & \text{The pixel (u,v) is a possible ZLP} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

2.2.3 Analysis of the change in characteristic vectors

Let I_0 be the image captured under zoom parameter $Z - \delta Z$, and I_1 be the image captured under zoom parameter Z. We cut out the capturing region of I_1 from I_0, and magnify the cut image to the image size of I_1. This is the reference image I'_0. Let $\boldsymbol{x}_1(u,v)$ be a characteristic vector of pixel (u,v) in I_1, and x'_0 be in I'_0.

The difference $\Delta\boldsymbol{x}(u,v)$ between $\boldsymbol{x}_1(u,v)$ and $\boldsymbol{x}'_0(u,v)$ is defined as follows:

$$\Delta\boldsymbol{x}(u,v) = \boldsymbol{x}_1(u,v) - \boldsymbol{x}'_0(u,v) \tag{2}$$

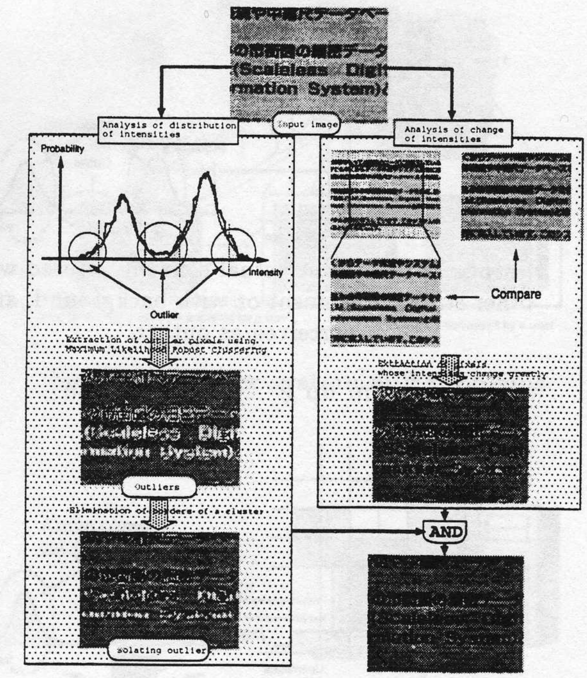


Figure 8: **Flow of extraction of ZLPs.**

We classify pixels into one cluster and outliers by probability density of $\Delta \boldsymbol{x}(u, v)$ using MARC. Outlier pixels are considered to be possible ZLPs.

As a weight concerning a possible ZLP through this analysis, I'_{uv} is defined.

$$I'_{uv} = \begin{cases} 1, & \text{The pixel (u,v) is a possible ZLP} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

2.2.4 Fusion of the results of two analyses

In the analysis of distribution of characteristic vectors, we can not classify pixels into clusters in stable according to the condition of an image. Pixels that are perceived as a principle cluster in a local area are perceived as an outlier in a whole image. On the other hand, in analysis of change in characteristic vectors, it is difficult to distinguish between the border of the structual element and ZLPs. In this study, we used both of these analyses together. We consider logical AND of possible ZLPs acquired by two analyses.

A weight $\tilde{I_{uv}}$ concerning ZLPs is described as follows.

$$\tilde{I_{uv}} = I_{uv} \cap I'_{uv} \quad (4)$$

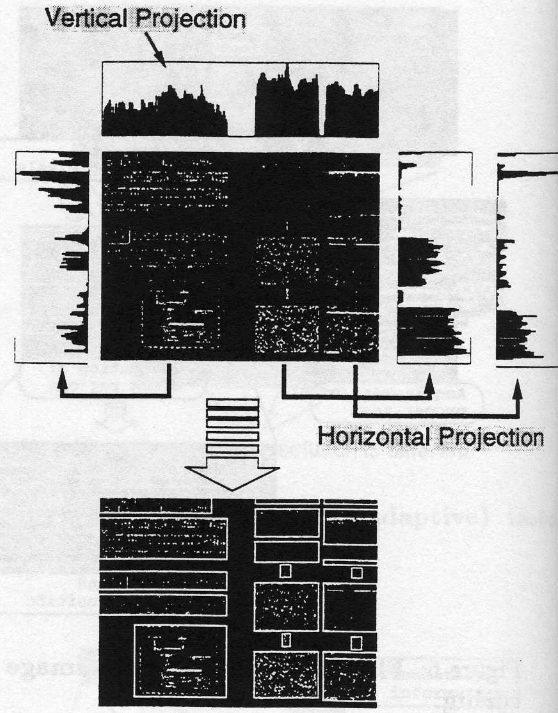


Figure 9: **Generation of block regions (white pixels show ZLP).**

Then, the region that has a lot of ZLPs should be zoomed up.

2.3 Generation of block regions

We classify ZLPs into some block regions by space distribution. Then a structure of a document can be known clearly, and we can do an effective presentation. Here, a technique of a layout analysis of a document image can be utilized. A block region is generated using both the horizontal and vertical projection distributions of ZLPs. It is divided by spot that histgram of projection distribution amounts to 0. In order to avoid minute division, a median filter is used in the projection distribution, and small division is eliminated. Each block is formed on suitable size using a maximum and a minimum value of coordinates of ZLPs which are included in its region.

Figure 9 shows one of the experimental results.

By the way, the technique which we use this time is assumed to be able to express a region by a rectangle. So, when a region is complicated, this technique does not sometimes act on stability. It also is not stable concerning an inclination of a document. To this, application of a technique of a complicated layout analysis is a subject to future[5][6].

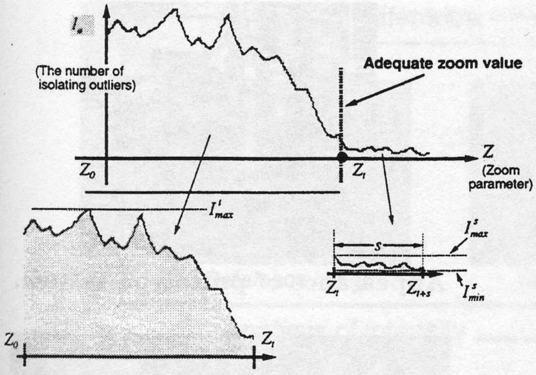


Figure 10: **Determination of appropriate zoom parameter.**

2.4 Determination of appropriate zoom parameters

It is supposed that the zoom value changes from low (wide) to high (tele) following time. About the image captured in the zoom value z_t by time t, $I_a^{z_t}$ is calculated by the following:

$$I_a^{z_t} = \sum_{u,v} \tilde{I_{uv}} \quad (5)$$

The value $I_a^{z_t}$ is counted on various zoom parameters. The change in $I_a^{z_t}$ with an increase in the zoom parameter is monitored. The zoom value with the stagnant starting time is selected as the appropriate parameter (Figure 10).

The stagnant starting time is decided as follows. We consider the number of ZLPs $[I_a^{z_t} \ldots I_a^{z_{t+s}}]$ during the time $[t \ldots t+s]$ when the time is t. Then we acquire the maximum value I_{max}^s and the minimum value I_{min}^s of $[I_a^{z_t} \ldots I_a^{z_{t+s}}]$. When a condition of

$$I_{max}^s - I_{min}^s < T \quad (6)$$

is satisfied, the time t is the stagnant starting time. The threshold T is decided as the following using the maximum value I_{max}^t of $[I_a^{z_0} \ldots I_a^{z_t}]$:

$$T = \gamma \cdot I_{max}^t \quad (7)$$

s, γ set 3, 0.1 as an experience.

One of experimental results is shown in Figure 11, 12, 13 and 14. Figure 11 shows a set of original images. Figure 12 shows a set of analyzed images. The white pixels are ZLPs. The change in the number of ZLPs with a change in the zoom parameter is shown in Figure 13, and Figure 14 shows a proper zoom image which is acquired from Figure 13.

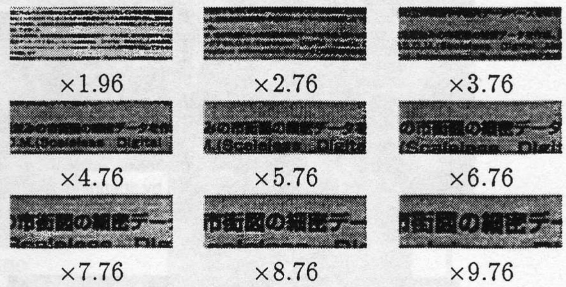


Figure 11: **A series of original images.**

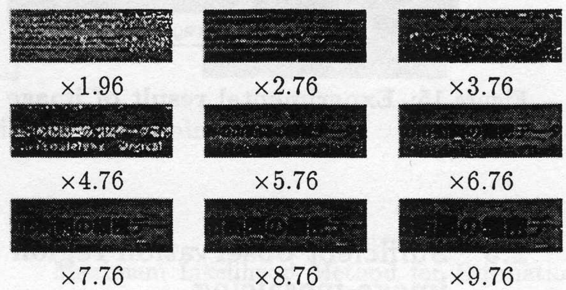


Figure 12: **A series of analyzed images (white pixels show ZLP).**

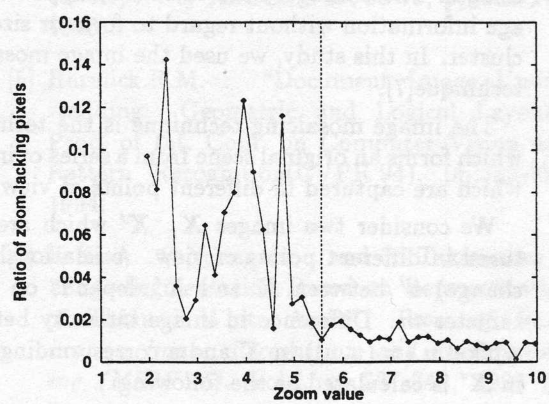


Figure 13: **Transition of the number of ZLPs.**

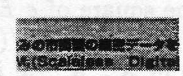


Figure 14: **Proper zoom image.**

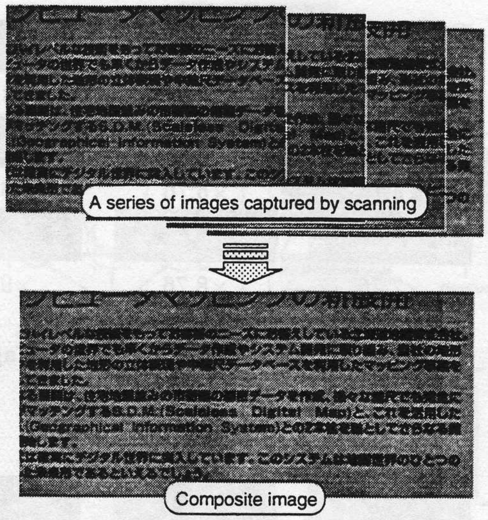


Figure 15: **Experimental result of image mosaicing.**

2.5 Sufficient observation region and image-mosaicing

In many cases, a sufficient observation region of a block can not be guaranteed with a single exposure by an adequate image resolution. Therefore, we need to control pan and tilt angles to capture images. These images need to be overlapped by image information without regard to form or size of a cluster. In this study, we used the image mosaicing technique[7].

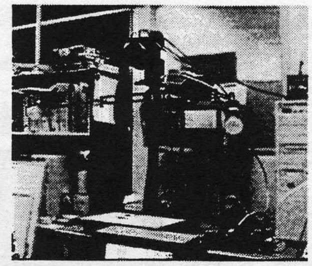
The image mosaicing technique is the technique which forms an original scene from a series of images which are captured in different points of view.

We consider two images $\boldsymbol{X}$, $\boldsymbol{X}'$ which are captured in different points of view. A relationship (a change) $\boldsymbol{u}'$ between $\boldsymbol{X}$ and $\boldsymbol{X}'$ depends on a parameter $\boldsymbol{m}$. Difference in image intensity between a pixel $\boldsymbol{u}$ $(= (u, v))$ in $\boldsymbol{X}$ and a corresponding pixel in $\boldsymbol{X}'$ is calculated as the following:

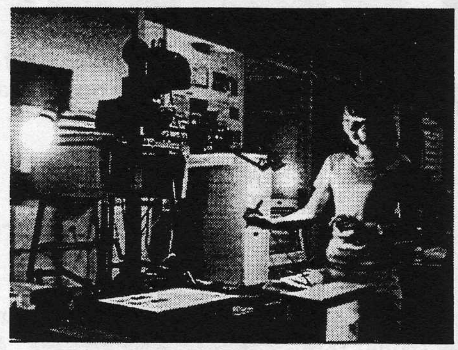
$$e(\boldsymbol{u}) \;=\; \boldsymbol{x}'(\boldsymbol{u} + \boldsymbol{u}'(\boldsymbol{m})) - \boldsymbol{x}(\boldsymbol{u}) \qquad (8)$$

We estimate the parameter $\boldsymbol{m}$ which minimizes the sum of the square of e $E = \sum_{\boldsymbol{u}} e^2$ using Hessian method. Here, $\boldsymbol{x}(\boldsymbol{u})$, $\boldsymbol{x}'(\boldsymbol{u})$ shows the intensity of the pixel $\boldsymbol{u}$ in $\boldsymbol{X}$, $\boldsymbol{X}'$. $\boldsymbol{u}'$ is described using perspective projection model.

One of the experimental results is shown in Figure 15.



Appearance of prototype system.



Circumstance using our system.

Figure 16: **Prototype system.**

3 Prototype system and experimental results

We constructed a prototype of an inhomogeneous image capturing system (Figure 16).

For the active camera, we use on EVI-D30 (SONY). Pan, tilt and zoom of the camera are controlled through a RS-232C interface from the HOST computer. Images are captured in the HOST computer using an image capturing board.

A user has to choose a block which is displayed. In our system, there is an interface which a user can choose a block using a laser pointer (Figure 17).

Figure 18 shows a captured inhomogeneous image of a document.

4 Conclusion

We have proposed a technique for camera-based inhomogeneous image capturing in order to develop an intelligent presentation system. Zoom parameters of the camera and the area to be focused are automatically determined. Images captured are merged using image mosaicing. We constructed a prototype of the inhomogeneous image capturing

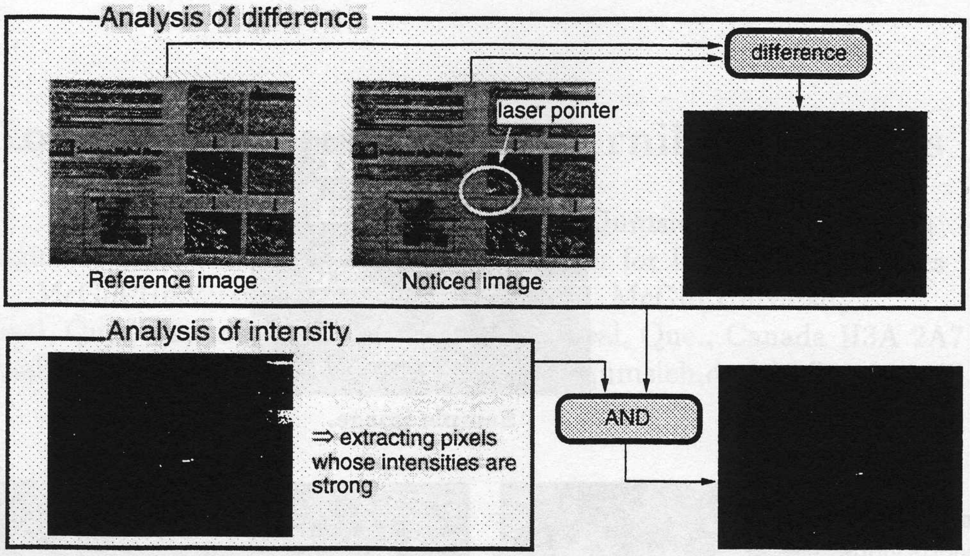


Figure 17: **Flow of detection of laser pointer.**

system using the proposed techniques. An experiment was conducted on live-images.

Our future works are:

- Linking between two images or an image and other information
- Structural or content analysis of presentation resources
- Recognition of an intention of a user
- Effective display method to viewers

Acknowledgement

This research was supported by a JSPS Research Fellowship for Young Scientists.

References

[1] M.Toda, T.Yamaguchi, T.Kawashima, and Y.Aoki : "Handling of Real Images for Intelligent Presentation System Using Active Camera", *Proc. of Int. Conf. on Information, Communication and Signal Processing (ICICS'97)*, Vol.II, pp.1093-1097, 1997.

[2] M.Toda, T.Kawashima, and Y.Aoki : "Active Sensing of Zooming on Vision", *Proc. of the 3rd Symposium on Sensing via Image Information (SII'97)*, pp.265-270, 1997 (in Japanese).

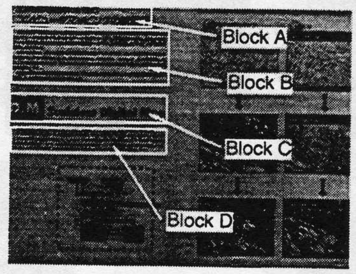
[3] N.Ichimura : "Robust Clustering Based on a Maximum Likelihood Method for Estimation of the Suitable Number of Clusters", *Technical Report of IEICE.*, PRU94-20, pp.9-16, 1994 (in Japanese).

[4] N.Ichimura : "Study on Inexhustive Region Segmentation by Robust Clustering", *Technical Report of IEICE.*, PRU95-78, pp.169-174, 1995.

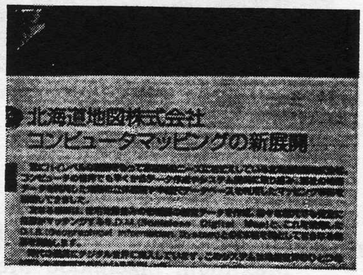
[5] Haralick.R.M. : "Document Image Understanding: Geometric and Logical Layout", Proc. of Int. Conf. on Computer Vision and Pattern Recognition(CVPR'94), pp.385-390, 1994.

[6] K.Kise, O.Yanagida, and S.Takamatsu : "Page Segmentation Based on Representation of Thinned White Areas", *Proc. of Meeting on Image Recognition and Understanding (MIRU'96)*, Vol.I, pp.337-342, 1996 (in Japanese).

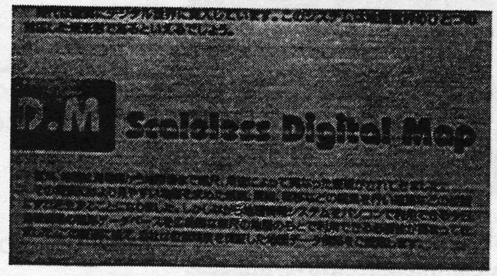
[7] R.Szeliski : "Image Mozaicing for Tele-Reality Applications", *Technical Report CRL*, 94/2, DEC Cambridge Research Lab., 1994.



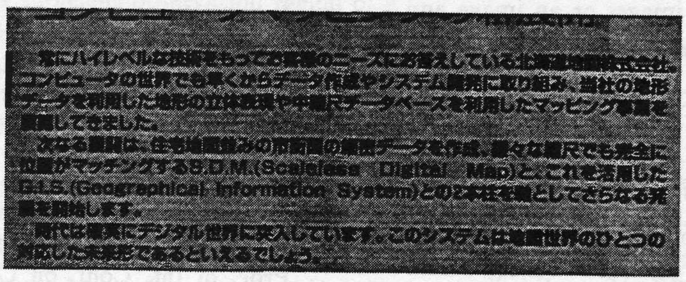
Sample image.



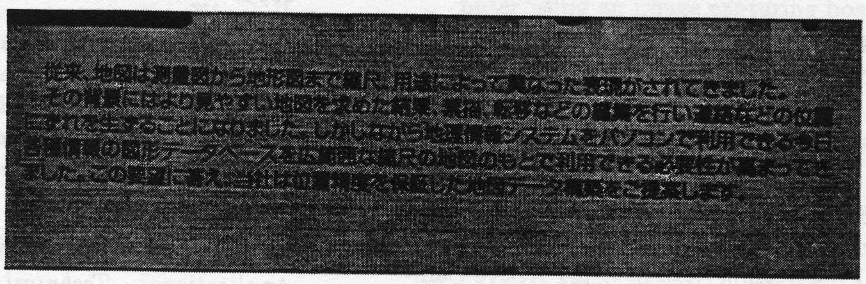
Block A.



Block C.



Block B.



Block D.

Figure 18: **Inhomogeneous image capturing using prototype system.**