MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11 - 13, 2002, Nara - ken New Public Hall, Nara, Japan

13—13

Robust Object Detection and Segmentation Based on Radial Reach Correlation

Yutaka Satoh, Hideki Tanahashi, Yoshinori Niwa*
HOIP, Softopia Japan / JST

Shun'ichi Kaneko†
Faculty of Engineering, Hokkaido University

Kazuhiko Yamamoto‡
Faculty of Engineering, Gifu University

Abstract

We propose a novel algorithm for robust object detection and segmentation, which is based on a new robust dissimilarity measure called as Radial Reach Correlation (RRC). The capability of detecting new objects from a complex background is one of the most important technologie for many vision systems. The usual approach for detecting new objects from a background is simple background subtraction. However, it is strongly influenced of brightness changes such as shadows and gain change of the imaging system. The RRC is a new robust dissimilarity measure and has a well-formed probabilistic model of binary or normal density. The RRC evaluates the local texture between a background image and the current scene and realize robust object detection under poor conditions. Experimental results with real images show the effectiveness of the proposed method.

1 Introduction

The capability of extracting new objects from a complex background is a fundamental and important technology for human tracking[1], motion analysis [2] and image sequence understanding [3, 4, 5]. The usual method for detecting objects from a complex background is simple background subtraction. However, simple background subtraction is susceptible to illumination change such as shadows, and when the brightness differences between objects and the background is small, it cannot detect the difference. These problems cause the consequent processes such as tracking, recognition, etc., to fail. In order to cope up with these problems, some algorithms such as color based subtraction technique [6] and the technique based on optical flows [7] have been proposed. However, it is well known that such techniques have problems in stability and calculation costs.

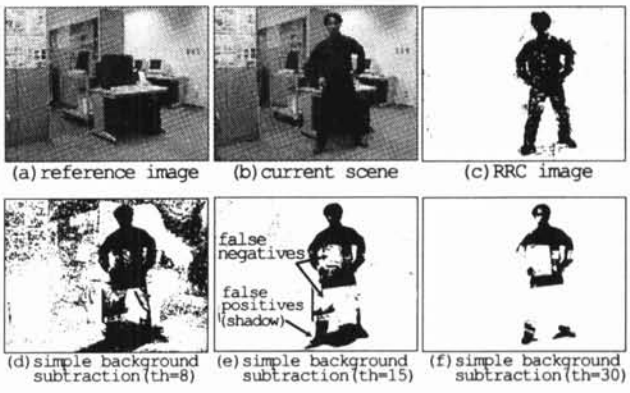


Figure 1: Overview.

In this paper, we propose a novel algorithm for robust object detection and segmentation, which is based on a new robust dissimilarity measure. This paper includes a theoretical modeling and analysis of the proposed method and some experimental results with real images.

2 Radial Reach Correlation (RRC)

Figure 1 shows the processing result of the RRC and the conventional technique. Figure 1 (d)-(f) shows a result of simple background subtraction which contains the following problems:

- Regions with the small brightness differences are undetectable (false negatives).
- Shadows are detected (false positives).

Figure 1 (c) shows the RRC image. This figure shows that such kind of problems are solved, and the person's region is detected well.

The RRC consists of the following two characteristic elements:

*Address: 4-1-7 Kagano, Ogaki City, Gifu 503-8569 Japan.pan. E-mail: satoh@hoip.jp

†Address: Kita-13 Nishi-8 Kita-ku, Sapporo 060-8628 Japan. E-mail: kanekos@coin.eng.hokudai.ac.jp

‡Address: 1-1 Yanagido, Gifu 501-1193 Japan. E-mail: yamamoto@info.gifu-u.ac.jp

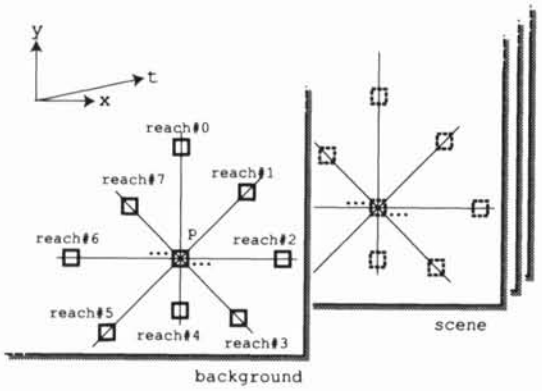


Figure 2: Radial Reach Correlation.

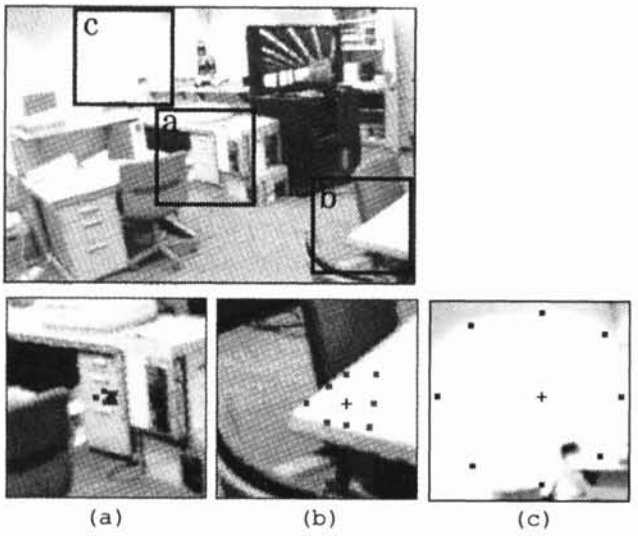


Figure 3: RRC-pairs.

(1) Search for the pixel with the brightness difference more than a threshold T_P from the current pixel (p). This operation is repeated about eight directions in the shape of radiation (see Figure 2), and 8 sets of the 'RRC pairs' are generated.

(2) The evaluation value of the current pixel (p) is obtained by the sign of brightness difference of each pairs which are expressed as a binary code (8bit).

(1) is the rule for obtaining the robustness to a noise. Robust pairs can be obtained by setting up T_P based on the noise characteristic of an imaging system (see 3.1 for further details). And in order to obtain isotropy and to introduce majority processing, we use pairs of eight directions (see Figure 2).

(2) is the rule for obtaining the robustness to illumination change such as shadows. In order to obtain the characteristic which does not detect an offset of brightness, only signs of brightness differences are used.

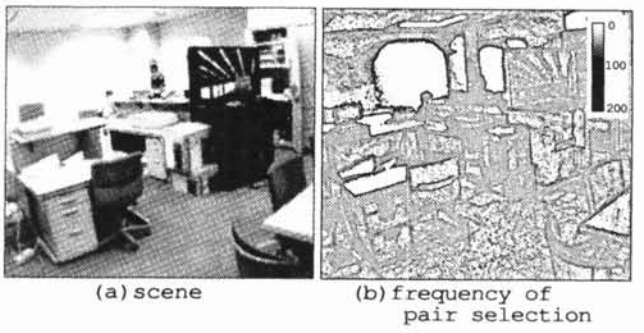


Figure 4: Distribution of frequency of pair selection.

In actual calculation, the RRC pairs are determined first. The arbitrary positions in the image are given by vector $\boldsymbol{p} = (x, y)^T$. The directions which lengthens the reach are given by the direction vectors $\boldsymbol{d}_k$ $(k = 0, 1, \cdots, 7)$. $\boldsymbol{d}_0 = (1, 0)^T$, $\boldsymbol{d}_1 = (1, 1)^T$, $\boldsymbol{d}_2 = (0, 1)^T$, $\boldsymbol{d}_3 = (-1, 1)^T$, $\boldsymbol{d}_4 = (-1, 0)^T$, $\boldsymbol{d}_5 = (-1, -1)^T$, $\boldsymbol{d}_6 = (0, -1)^T$, $\boldsymbol{d}_7 = (1, -1)^T$. The $\{r_k\}_{k=0}^7$ which satisfies the following expression is calculated.

$$r_k = \min\{r \Big| |f(\boldsymbol{p} + r\boldsymbol{d}_k) - f(\boldsymbol{p})| \geq T_P(\boldsymbol{p})\}, \qquad (1)$$

where f is the background image, and $T_P(\boldsymbol{p})$ is the threshold of the brightness difference (see 3.1 for further details about $T_P(\boldsymbol{p})$). The set of $f(\boldsymbol{p})$ and $f(\boldsymbol{p}+r\boldsymbol{d}_k)$ with which are satisfied of the expression (1) is called 'RRC Pair'. Figure 3 shows an example of the RRC-pairs generated with a real image. The RRC-pairs with long reach are generated in (c) by which brightness is saturated and the texture is lost. On the other hand, the RRC-pairs with short reach is generated in (a) with a fine texture. Figure 4 shows the distribution of frequency of pair selection. In the region in which brightness is saturated under the influence of lighting, it turns out that the pair point is concentrating around it. Also in the region on a desk with a weak texture, it turns out that the pair point is concentrating on the surrounding edge portion. However, in the region with a dense texture, the pair point is distributed uniformly.

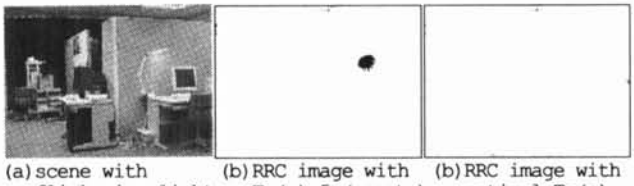
Next, the RRC evaluation value is calculated using the RRC Pair.

$$\begin{aligned} b_0(\boldsymbol{p}) &= \begin{cases} 1 & (\text{if } f(\boldsymbol{p} + \boldsymbol{d}_0 r_0) \geq f(\boldsymbol{p})) \\ 0 & (\text{otherwise}) \end{cases} \\ &\vdots \qquad (2) \\ b_7(\boldsymbol{p}) &= \begin{cases} 1 & (\text{if } f(\boldsymbol{p} + \boldsymbol{d}_7 r_7) \geq f(\boldsymbol{p})) \\ 0 & (\text{otherwise}) \end{cases} . \end{aligned}$$

Let $b_0(\boldsymbol{p})\sim b_7(\boldsymbol{p})$ (8bit) be the evaluation value of the interest pixel of the background image. The evaluation value $b'_0(\boldsymbol{p})\sim b'_7(\boldsymbol{p})$ of the current scene ($g(\boldsymbol{p})$) is calculated in the same procedure.

(a) $T_P(p)=0$ (const.) (b) $T_P(p)=4$ (const.) (c) $T_P(p)=8$ (const.)

Figure 5: Settings of T_P.



(a) scene with flickering light (b) RRC image with $T_P(p)=5$ (const.) (b) RRC image with optimal $T_P(p)$

Figure 6: Scene with flickering light.

Next, in order to evaluate the similarity of the RRC evaluation value between the background image and the current scene, B is defined as follows:

$$B(\boldsymbol{p}) = \sum_{k=0}^{7} d_k(\boldsymbol{p}), \quad (3)$$

$$d_k(\boldsymbol{p}) = \{b_k(\boldsymbol{p}) \cdot b'_k(\boldsymbol{p}) + \overline{b_k(\boldsymbol{p})} \cdot \overline{b'_k(\boldsymbol{p})}\}, \quad (4)$$

where $\overline{x} = 1 - x$. Then, the RRC image $R_{(\boldsymbol{p})}$ is obtained by comparing $B(\boldsymbol{p})$ with judgment threshold T_B (see 3.2 for further details).

$$C(\boldsymbol{p}) = \begin{cases} 1 & (B(\boldsymbol{p}) > T_B) \\ 0 & (\text{otherwise}) \end{cases}. \quad (5)$$

Figure 1 (c) shows the RRC image. The pixels with $C(\boldsymbol{p}) = 1$ are shown as white pixels and the others ($C(\boldsymbol{p}) = 0$) are shown as black pixels. This figure shows that the human region has detected well.

3 Design of thresholds

3.1 Brightness difference threshold : $T_P(p)$

$T_P(\boldsymbol{p})$ specifies the minimum brightness difference in the RRC pair. Figure 5 shows the examples of the RRC images with the constant T_P (0, 4, 8) for all pixels. In the case of $T_P(\boldsymbol{p}) = 0$ (Figure 5 (a)), false positives have occurred in the region with weak texture because the RRC pairs are generated regardless of the brightness differences. If the brightness difference is small, it will be easy to be influenced of a noise. On the other hand, in Figure 5 (c), the influence of noises are reduced by suitable reach length being set up according to the texture conditions.

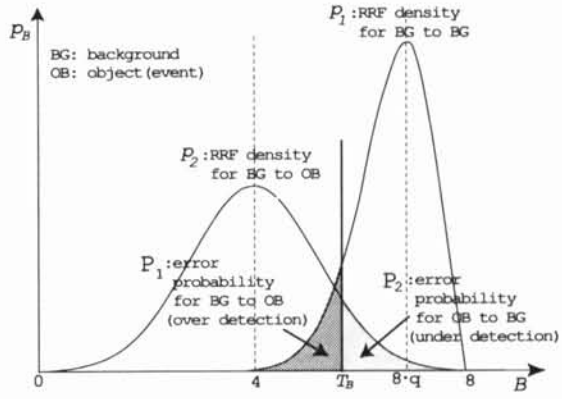


Figure 7: Probability distributions of p_1, p_2.

As a simple way, constant value can be used as the value of $T_P(\boldsymbol{p})$ for all $\boldsymbol{p}$ as mentioned above. However, usually T_P can be rationally determined according to the noise characteristic of an imaging system. $T_P(\boldsymbol{p})$ should be set as 2σ of the distribution of the measured noise histogram of the position $\boldsymbol{p}$. Theoretically, the RRC pair is not influenced about 95% of the noise by this setup.

Figure 6 (a) shows a scene with a flickering light. Supposing T_P is constant at all pixels, false positives will occur in the flickering regions (Figure 6 (b)). In Figure 6 (c), since optimal T_P is automatically given for every pixel, false positives are not seen.

3.2 Judgment threshold : T_B

T_B is the threshold for judging whether the pixel $\boldsymbol{p}$ is similar or not between current scene and background image. We can utilize the statistical decision theory [9] to design T_B. Expressions (3) and (4) judge match between $b_k(\boldsymbol{p})$ and $b'_k(\boldsymbol{p})$. B takes the value of 0 to 8 according to the number with which $b_k(\boldsymbol{p})$ and $b'_k(\boldsymbol{p})$ matched about eight directions. Because the probability or frequency that any $d_k(\boldsymbol{p})$ equals 1 is expected to be q, the probability density of $B(0 \leq B \leq 8)$ is derived as follows:

$$p_B(B) = \sum_{u=0}^{8} p_u \delta(B - u) \quad (6)$$

$$p_u = \binom{8}{u} q^u (1-q)^{8-u}, \quad (7)$$

$$(u = 0, 1, \cdots, 8)$$

where $\delta()$ is the Delta or impulse function [8]. In the image sequence from a fixed view point, background regions are ideally expected to be similar, and when no noise, all $d_k(\boldsymbol{p})$ should be 1, resulting $B = 8$. In real data, however, as shown in Figure 7, p_B is expected to have the profile p_1 as the one for similar

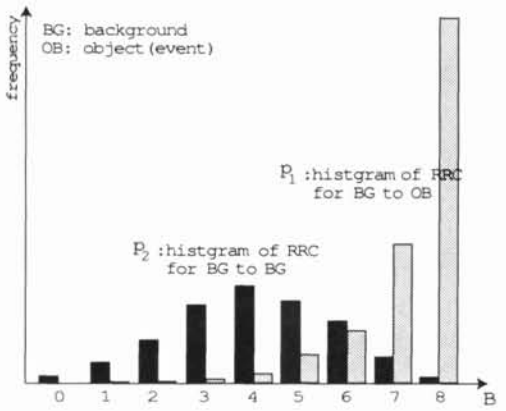


Figure 8: Actual measurement of p_1, p_2.

image pair, where the profile of distribution of B is depicted as an analog function although B is really a discrete quantity. Otherwise, a frame of the sequence including events should have the corresponding region having B around 4 ($q = 0.5$) as uncorrelated portions to the background image. This profile corresponding to this case can be represented by p_2 in Figure 7. The threshold T_B should be set up to discriminate these two peaks as shown in the figure. Then two kinds of error probability can be considered as follows: the first kind of error $P_1 = \int_0^{T_B} p_1(B)dB$ representing the error such that background pixels might be recognized as objects (over detection), while the second one $P_2 = \int_{T_B}^{8} p_2(B)dB$ corresponding to the case such that objects can be misled as background pixels (under detection). The loss function derived from these error probabilities can be calculated as the averaged risks $c_1 \cdot P_1 + c_2 \cdot P_2$, where c_1 and c_2 are costs for the respective mistakes. We can minimize the loss function according to $\partial(c_1 P_1 + c_2 P_2)/\partial T_B = 0$ for obtaining the optimal threshold T_B. The following condition can be derived.

$$\frac{p_1(T_B)}{p_2(T_B)} = \frac{c_2}{c_1}. \tag{8}$$

In the case uniform cost $c_1 = c_2$, we have the simpler version of the equation $p_1(T_B) = p_2(T_B)$. Figure 8 shows the example histogram corresponding to the densities p_1 and p_2, which were actually observed using real images shown in Figure 1 (a),(b). These show the feasibility of our approximation of the RRC by a binary probability distribution. Figure 9 shows the RRC images obtained by changing T_B. Since the distribution of p_1 and p_2 has overlapped, over detection and under detection are seen depending on a setup of the value. In this example, it is desirable to set $T_B = 6$.

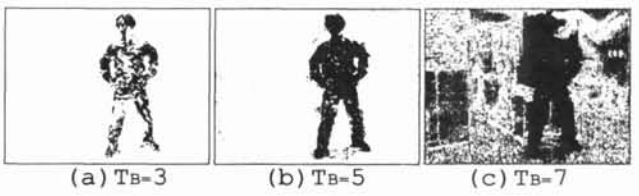


Figure 9: Tuning of T_B.

4 Experimental results

The experimental result under the normal conditions was already shown in Figure 1. Figure 10 shows experimental results under poor conditions. The background image was obtained where a window blind was closed. Figure 10 (a),(d) were obtained at the locations which receive direct sunlight. Figure 10 (b),(e) show the results of simple background subtraction.These figures show that detection has failed owing to the brightness change. On the other hand, although many noises are included, a person's domain can be viewed in the RRC images (Figure (c),(f)). The following reasons can be considered as cause of the noises.

(1) Reversal of the sign of brightness difference in the regions containing highlights or shadows.

(2) Saturation of the brightness in gloss parts or strong shadow parts.

When the influence of these false positives cannot be disregarded, updating of the background image is needed. However, updating the background image is a very difficult problem. Since the RRC is robust to brightness change, the frequency of updating the background image can be reduced.

Next, we describe about the experiment that used the stereo omni-directional system (SOS). The SOS (Figure 11) developed at our laboratory [10] provides real time omni directional images, stereo images and range information of surroundings in a sphere view. Figure 12 shows the mosaic panorama image obtained from the SOS. The parameter $T_B = 6$ was given and the RRC image (c) was obtained. The persons are detected well not related to their clothes and the background condition. Figure 12 (d) shows the extracted regions which obtained by Figure 12 (c) with dilation-erosion process.

Figure 13 zooms in near a portion shown with the rectangle in a Figure 12 (b). As for simple background subtraction, the region of a breast has not been detected because of its brightness distributions (Figure 13 (c)). Furthermore, shadow is strongly detected. On the other hand, the RRC image has detected the person's region well (Figure 13 (b)). And there are almost no influences by the shadows.

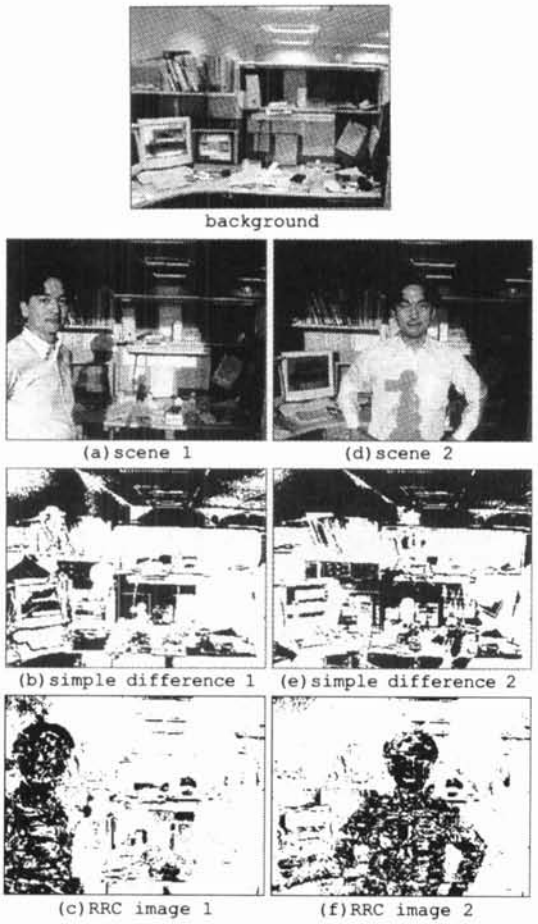


Figure 10: Detection of objects under poor illumination.

Next, the experiment which measures the height of a person who walks along 1 to 1.5m distance from the SOS was conducted using the distance information obtained from the SOS. Figure 14 shows the result of the experiment. The dashed line shows the true value. The result that used the RRC is stable. On the other hand, the result that used the simple background subtraction includes the substantial margin of error. These errors are caused by false negatives like shown in Figure 13 (c).

The calculation time per an image (900×298 pixels) with Pentium4 2.2GHz PC was as follows: Generation of the reach: around 0.6 (s). Generation of a RRC image: around 1 (ms). Although calculation cost of the generation of the reach is large, it is performed only once first. Therefore, the application to the real time processing can be considered.

5 Conclusions

We have proposed a novel algorithm for robust object detection called as Radial Reach Correlation (RRC), which is based on a new robust dissimilarity measure. Theoretical consideration and the experimental results showed the effectiveness of the proposal method. As future work, we would like to design a method for updating the background image.

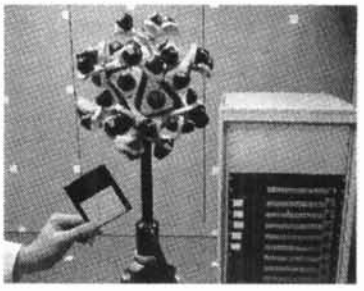


Figure 11: Stereo Omni-directional System (SOS).

References

[1] C.Wren, A.Azarbayejani, T.Darrell, A.Pentland, "Pfinder: Real-Time Tracking of the Human Body", IEEE Trans. PAMI, Vol.19, No.7, pp.780-785, 1997.

[2] J.K.Aggarwal and R.O.Duda, "Computer Analysis or Moving Polygonal Images", IEEE Trans. Computer, vol.C-24, pp.966-976, 1975.

[3] H.H.Nagel, "Analysis Techniques for image sequence", Proc. on 4th IJCPR, pp.186-211, 1978.

[4] R.Polana and R.Nelson, "Low Level Recognition of Human Motion", Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects, pp.77-82, 1994.

[5] I.Haritaoglu, D.Harwood, and L.Davis, "W4: Who? When? Where? What? A Real Time System for Detecting and Tracking People", Proc. the 3rd International Conference on Automatic Face and Gesture, pp.222-227, 1998.

[6] C.Stauffer and W.E.L.Grimson, "Adaptive background mixture models for real-time tracking", Proc. IEEE CVPR'99, Vol.II, pp. 246-253, 1999.

[7] Y. Mae, Y. Shirai, J. Miura and Y. Kuno, "Object Tracking in Cluttered Background Based on Optical Flow and Edges", Proc. IAPR ICPR96, Vol.1, pp. 196-200, 1996.

[8] A.Papoulis, Probability, Random Variables, and Stochatic Processes, McGraw-Hill Kogakusha, Tokyo, 1965.

[9] R.O.Dura and P.E.Hart, "Pattern Classification and Scene Analysis", John Wiley & Sons, New York, 1973.

[10] H.Tanahashi, D.Shimada, K.Yamamoto and Y.Niwa, "Acquisition of Three-Dimensional Information in a Real Environment by Using the Stereo Omni-directional System(SOS)", Proc. 3rd 3DIM, pp.365-371, 2001.

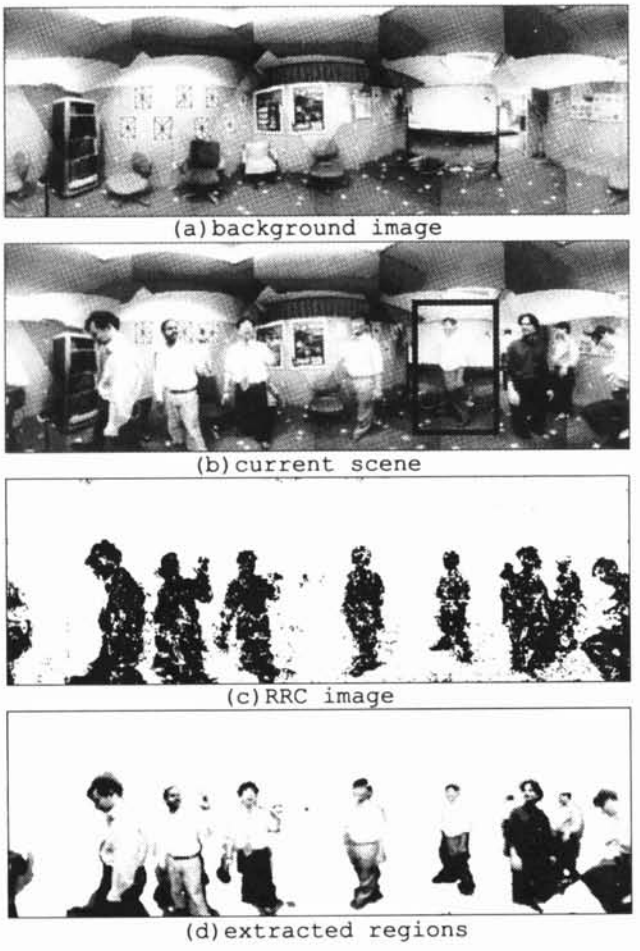


Figure 12: Experiment with panorama image.

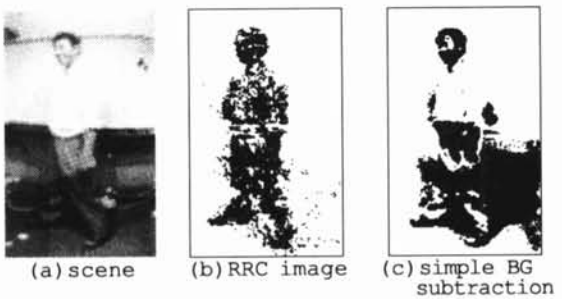


Figure 13: Predominancy of RRC.

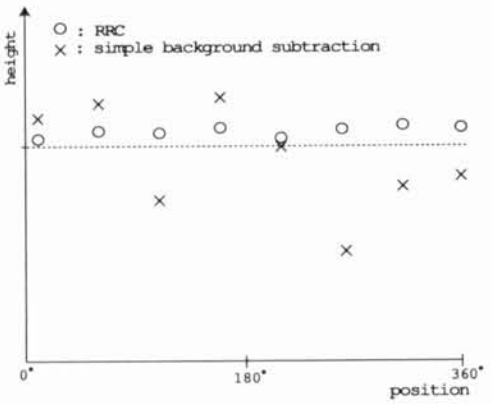


Figure 14: Measurement of person's height.