Surface Area Analysis for People Number Estimation

Hiroyuki Arai, Naoki Ito and Yukinobu Taniguchi
NTT Media Intelligence Laboratories, NTT Corporation, 1-1 Hikari-no-oka, Yokosuka, Kanagawa, Japan

Keywords: Projective Geometry, Camera Calibration, Surface Area, People Number Estimation, Congestion Monitoring.

Abstract: An important property of surface areas of objects as observed by a calibrated monocular camera is introduced; also improved techniques to apply the property to people number estimation are proposed. Standard surface area (SSA) is defined as the surface area of the reverse projection of an image-pixel onto a plane at specific height in the real world. SSA is calculated for each pixel according to camera calibration parameters. When the target object is bound to a certain plane, for example the floor plane, the sum of SSA along with the foreground pixels of one target object becomes constant. Therefore, simple foreground detection and SSA summation yield the number of target objects. This basic idea was proposed in a prior article, but there were two major limitations. One is that the original model could not be applied to the area directly below the camera. The other is that the silhouette of the target object was limited to a simple rectangle. In this paper we propose improved techniques that remove the limitations. Slant silhouette analysis removes the first limitation, and silhouette decomposition the second. The validity and the effectiveness of the techniques are confirmed by experiments.

1 INTRODUCTION

Crowd estimation or crowd level monitoring is important, especially in the fields of safety management and marketing. Crowd levels at a train station should be carefully monitored because crowded situations may cause an accident, for example a fall. In the field of marketing, audience acceptance of digital-signage is becoming more and more important. The number of people who remain in front of the screen is thought to be a key audience rating metric.

Existing image processing techniques to estimate the number of people are roughly divided into 3 types. The first covers “shape-detection based” techniques, for example, detection of human body, upper-body or head region(Sheng-Fuu Lin and Chao, 2001)(Min et al., 2008). The second consists of “tracking based” techniques (Rabaud and Belongie, 2006), (Zhao et al., 2007),(Sidla et al., 2006),(Antonini and Thiran, 2006), for example, tracking people by trajectory clustering in the spatio-temporal domain. The last, “feature-based” techniques, relate image features to the number of people by case-learning (Marana et al., 1998),(Cho et al., 1999),(Kong et al., 2006),(Wen et al., 2011).

Crowd estimation schemes based on image processing are expected to meet the following requirements, especially for the application of audience estimation. 1)Must be stable even if crowds are present. 2)Must have low computation cost. 3)Must be easy to set up (because many systems will be needed). If crowds are present, shape-detection becomes unstable because many confusing shapes appear in each image, and tracking also becomes unstable because too many feature points appear around each other. Fortunately, feature-based techniques are more stable than the other two, because the basic operations used to get the feature vector, such as edge detection or foreground detection, can be done even if crowds are present. Feature-based techniques also satisfy the computation cost requirement. However, existing feature-based techniques demand a large amount of ground-truth data. Given the very large number of cameras needed, preparing enough data would take far too long and be too expensive.

To satisfy all the above requirements, an a priori algorithm that can estimate the number of people by analyzing foreground images was proposed(Arai et al., 2009). The basic idea is that, after the camera is calibrated, a rough estimation of the number of people can be made by determining the number of foreground pixels and their positions in the image. To realize this idea, the concept of ”standard surface area (SSA)” which quantitatively represents the extension of each pixel in the real world was introduced. The prior article showed that the sum of SSA in the case of

Arai H., Ito N. and Taniguchi Y..
Surface Area Analysis for People Number Estimation.
DOI: 10.5220/0004687000860093
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 86-93
ISBN: 978-989-758-004-8

a rectangle silhouette standing on the floor becomes invariant. By using this property of SSA and an occlusion model that estimates the influence of occlusion on foreground image, the number of people can be estimated from foreground images.

According to this technique, a system for measuring the advertizing effectiveness of digital signage was developed. However, while in operating the system, several problems were revealed. The original technique, implemented in the system, has two important limitations. One is that it can not be applied to the area directly below the camera. The other is that the silhouette of the target object is limited to a simple rectangle. The first limitation is especially serious, because many cameras will be fixed to the ceiling to cover the areas underneath them. Although the second limitation is not so serious in the case of people number estimation, because the silhouette of standing or walking people can be roughly approximated to the rectangle model, the ability to handle arbitrary silhouettes is expected to improve measurement precision, and lead to new applications.

In this paper, we extend the theory proposed in prior article and introduce improved techniques, slant silhouette analysis and silhouette decomposition, to be able to realize directly downward capture and arbitrary silhouettes.

Section 2 provides the theoretical background of the method. The slant silhouette analysis is detailed in Section 3. Section 4 describes the silhouette decomposition technique. Experiments to confirm the validity of the theory and the effectiveness of the estimation technique are shown in Section 5. Our conclusion is given in Section 6.

2 FUNDAMENTAL THEORY OF NUMBER ESTIMATION

This section describes the basic theory of the geometrical properties of SSA and the concept of number estimation. Here, some part of the theory has been modified (simplified) from the former article().

2.1 Preconditions

The algorithm, described below, is valid given the following assumptions,

(a) camera has been calibrated,

(b) target object, for example human body, can be approximated as a rigid silhouette of known size.

(c) image resolution is high enough so that each target object occupies many pixels in the image

(d) target objects are randomly positioned on the floor.

2.2 Basic Idea for Number Estimation

The basic idea is to quantitatively relate each foreground pixel to some part of the surface area of the target object. To simplify the explanation of this idea, we consider the case of one person standing at known position (X,Y) on the floor and assume that the human body can be approximated as a rectangle silhouette with known size, see Figure 1.

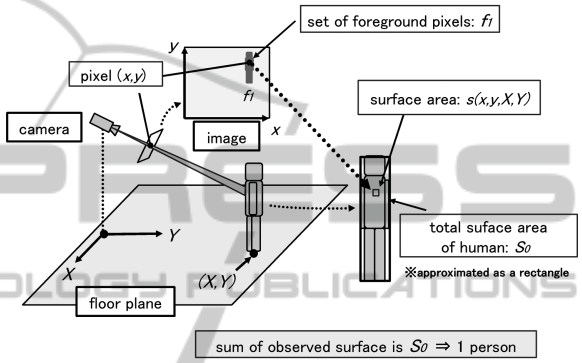

Figure 1: Basic idea: correspondence of pixel to partial surface area of human body.

Let $s(x,y,X,Y)$ denote the partial surface area of the rectangle(person) standing at (X,Y) on the floor, that corresponds to pixel(x,y), f_1 denotes the total set of foreground pixels that correspond to the rectangle, and S_0 represents the entire surface area of a single rectangle silhouette in the real world; the relationship between them can be written as follows,

$$\sum_{(x,y)\in f_1} s(x,y,X,Y) = S_0. \tag{1}$$

$s(x,y,X,Y)$ can be calculated by analyzing the reverse projection of pixel (x,y) onto the rectangle standing at (X,Y) on the floor. If the person is perfectly detected as foreground in the image, the sum of $s(x,y,X,Y)/S_0$ along with the foreground pixels becomes 1, because the sum of $s(x,y,X,Y)$ approaches S_0 which means the entire surface of the rectangle in the real world. Therefore, pixel(x,y) brings proof of $s(x,y,X,Y)/S_0$ for one person's existence at (X,Y). By simple extension, the number of people N can be estimated as follows,

$$N = \frac{\sum_{(x,y)\in F} s(x,y,X,Y)}{S_0}, \tag{2}$$

here, F is the set of foreground pixels for all people in the image, and S_0 is the surface area, in the real world measure, of the silhouette of a single person.

2.3 Definition of Standard Surface Area

The assumption that the positions of the people (X,Y) are known is not realistic, especially in the case of crowded scenes. As the position of the person changes, the corresponding surface area in the real world changes, see Figure 2.

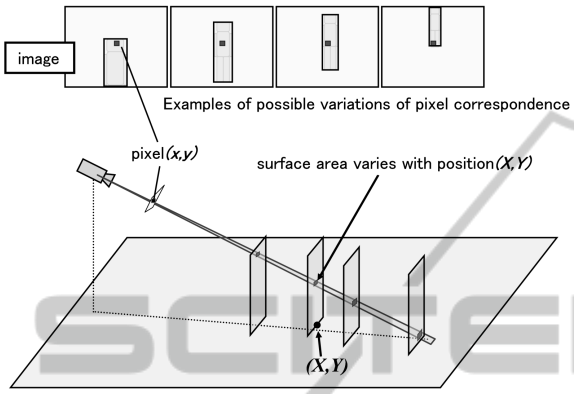

Figure 2: Variation in pixel correspondence to surface area in real world.

Therefore, equation(2) cannot be applied naively. To handle position variation, the concept of "standard surface area:SSA" is introduced. Pixel (x,y) can correspond to any of the many positions at which a person could stand, and its surface area directly depends on the height of the projection on the human body. SSA is defined as follows,

$$\hat{s}(x,y) = s(x,y,X_\eta,Y_\eta), \quad \{X_\eta,Y_\eta \mid \eta = Z\{(x,y),(X_\eta,Y_\eta)\}\}, \tag{3}$$

here $Z\{(x,y),(X,Y)\}$ is a function that calculates the Z-position(height) of the projection of pixel (x,y) at position (X,Y) on the floor. η is a fixed parameter, but can be arbitrary determined within the range of $\eta < Tz$, where T_z is the height of the camera above the floor. SSA: $\hat{s}(x,y)$ gives a quantitative measure of the extension of each pixel, in the real world, and plays an important role in solving the problem of position variation, as mentioned in the following section.

In the prior article(Arai et al., 2009), $\hat{s}(x,y)$ is defined in another way, that is, the average surface area of the projection of pixel(x,y) from $(Z=0)$ to $(Z=H)$, here H is known height of the target object. Details are not shown here, but the height that yields the average surface area is the same at any pixel in an image. Therefore, the former definition can be considered as a special case of the new definition.

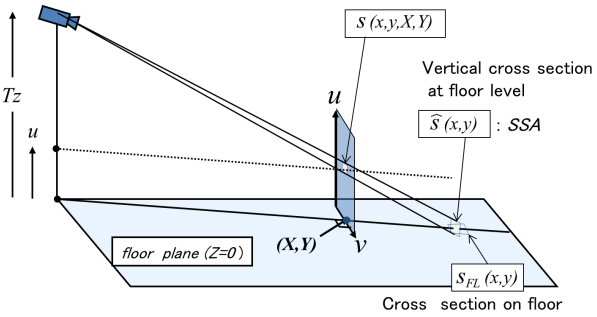

Figure 3: Geometrical configuration of SSA (case $\eta = 0$).

2.4 Invariance of Sum of SSA (Case of Upright Rectangle Model)

Here, let us consider the sum of $\hat{s}_{x,y}$ along with the foreground pixels for one person similar to equation(1). To simplify the problem, we consider the case of $\eta = 0$ and approximate the silhouette of the target object as a rectangle with height H and width D, see Figure 3. Assumption (c) in 2.1 indicates that pixel size is very small(image resolution is very high), so the sum of $\hat{s}(x,y)$ can be approximated as the following integration formula

$$\sum_{(x,y)\in f_1} \hat{s}(x,y) = \iint_{F_1} \hat{s}(x,y)dxdy. \tag{4}$$

Here, the range of this integration, F_1, is the continuous region that corresponds to discrete region f_1 on the image plane. As shown in Figure 3, considering a rectangle standing front-on to the camera, with vertical and horizontal edges (u,v), the following equation,

$$\frac{dudv}{dxdy} = s(x,y,X,Y) \tag{5}$$

is derived from the definition of $s(x,y,X,Y)$ as a transformation coefficient from image area to surface area in the real world. By using equation(5), equation(4) can be rewritten as follows,

$$\sum_{(x,y)\in f_1} \hat{s}(x,y) = \iint_{F_1'} \frac{\hat{s}(x,y)}{s(x,y,X,Y)}dudv. \tag{6}$$

Here, the range of this integration F_1' is the entire surface of the rectangle. The integrand of equation(6) can be easily derived from the proportionality of the cross section and height of cone, as follows,

$$\frac{\hat{s}(x,y)}{s(x,y,X,Y)} = \left(\frac{Tz}{Tz-u}\right)^2. \tag{7}$$

Therefore, equation(6) can be rewritten as follows,

$$\begin{aligned}\sum_{(x,y)\in f_1}\hat{s}(x,y) &= \iint_{F_1'}\frac{Tz^2}{(T_z-u)^2}dudv \\ &= \int_0^D\int_0^H\frac{Tz^2}{(T_z-u)^2}dudv \\ &= S_0(1+\frac{T_zH}{(Tz-H)Tz}) \\ &= S_{unit}. \end{aligned} \quad (8)$$

As can be seen in equation(8), the sum of $\hat{s}(x,y)$ along with f_1, the set of the foreground pixels of each rectangle, does not depend on (x,y,X,Y), only on Tz:the vertical position of camera, H:the height of rectangle, and S_0:the total surface area of the rectangle in the real world. We denote this invariant S_{unit}.

By simple analogy to equation(2), the number of people can be estimated by using following equation,

$$N = \frac{\sum_{(x,y)\in F}\hat{s}(x,y)}{S_{unit}}, \quad (9)$$

here, F is the set of foreground pixels for all people in the image. This enables people number to be estimated without knowing the position of each person.

While there are several ways to calculate $\hat{s}(x,y)$ in practice(Arai et al., 2009), we show here a simple approach. Considering reverse projection of 4 points, $(\{x\pm1/2\},\{y\pm1/2\})$, onto floor plane, $s_{FL}(x,y)$; the cross section on floor plane can be calculated (see Figure 3). The angle between pixel (x,y) and floor plane, $\alpha(x,y)$, can be easily calculated. Using these values, the $\hat{s}(x,y)$ for each pixel can be calculated as follows,

$$\hat{s}(x,y) = s_{FL}(x,y)\cdot tan[\alpha(x,y)]. \quad (10)$$

2.5 Problems of Upright Rectangle Model

The former method, which is based on the upright rectangle model, has two problems. One is that it can not be applied to areas directly below the camera. On viewing an upright rectangle directly from above, the silhouette becomes a thin line or disappears. In such cases, $\hat{s}(x,y)$ goes to a significantly large value or to infinity. Therefore, number estimation fails for areas directly below the camera. The other is that the silhouette of target object is limited to a simple rectangle. These problems reduce the measurement accuracy, and also narrow the application, as mentioned in section 2.

3 SLANT SILHOUETTE ANALYSIS

In this section, we extend the original theory and technique to be able to deal with the directly downward case. We introduce the slant silhouette model to solve this problem.

3.1 Basic Idea of Slant Rectangle Model

The fundamental problem of the upright rectangle model mentioned above is that the silhouette almost disappears with downward facing cameras(pixels). Therefore, the geometrical model should be changed to have non-zero area in the case of downward observation. Considering that the target is people, a cuboid or ellipsoid body might be the first choice. However, analytical calculation of the silhouette of an ellipsoid body is significantly difficult with perspective projection. Cuboids raise another problem; because a cuboid has more than one plane, analytical calculation will be almost impossible.

We introduce the slant rectangle model to solve this problem. Slant rectangles, for example in Figure 4, can well approximate cuboids, in the sense of silhouette approximation. They have non-zero silhouettes in plan and because they have only one plane, analytical calculation is not so difficult.

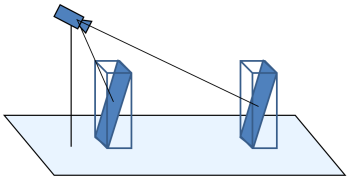

Figure 4: Slant rectangle model (approximation of cuboid).

3.2 SSA of Slant Rectangle Model

Figure 5 shows the geometrical configuration of the slant rectangle model. The rectangle is inclined backward at a specific angle β. The top of the rectangle is at height H, so the long edge has length of $H/cos\beta$. β should be determined to yield an area of the same order as the target object in the case of downward observation.

In this model, SSA: $\hat{s}(x,y)$ is defined as the cross section of pixel (x,y) on the β-slant plane at floor level(see Figure 5). Considering the ratio of $s_{FL}(x,y)$ and $\hat{s}(x,y)$, and applying sine theorem, yields

$$\hat{s}(x,y) = \frac{sin[\alpha(x,y)]}{cos[\alpha(x,y)-\beta]}s_{FL}(x,y). \quad (11)$$

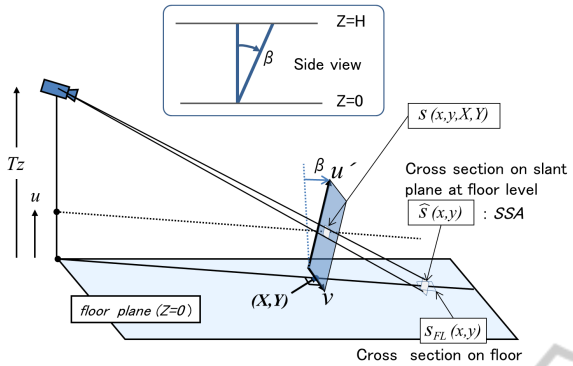

Figure 5: SSA for slant rectangle case.

Next, let us consider the property of SSA in this model. The $s(x,y,X,Y)$ is cross section of pixel (x,y) on the slant rectangle which stands at (X,Y) on floor plane. Suppose that this cross section is at the height of u. A consideration of the coordinate system on the slant rectangle (u',v) , as in Figure 5, yields

$$\frac{\hat{s}(x,y)}{s(x,y,X,Y)} = \left(\frac{Tz}{T_z-u}\right)^2, \quad (12)$$

$$s(x,y,X,Y) = \frac{du'dv}{dxdy}, \quad (13)$$

$$u = u'cos(\beta). \quad (14)$$

Considering the sum of $\hat{s}_{x,y}$ along with foreground pixels of one object(person) yields

$$\sum_{(x,y)\in f_1} \hat{s}(x,y) = \frac{1}{cos(\beta)} \int_0^D \int_0^H \left(\frac{T_z}{Tz-u}\right)^2 dudv$$

$$= \frac{1}{cos(\beta)} S_{unit}. \quad (15)$$

Here, S_{unit} is the same as in equation(8). Therefore, in the case of the slant rectangle model, SSA, $\hat{s}(x,y)$, is calculated by equation(11), and the estimated number is calculated by

$$N = \frac{\sum_{(x,y)\in F} \hat{s}(x,y)}{S_{unit}} cos(\beta). \quad (16)$$

By using the slant rectangle model, the silhouette has a non-negligible area even directly below the camera. As you can see in equation(11), SSA: $\hat{s}(x,y)$ does not diverge, because $(\alpha-\beta)$ can not reach $\pi/2$. This model thus solves the problem of observation directly under the camera.

In addition, if the slant angle β is 0, all equations of the slant model devolve into those of the upright model mentioned before.

4 EXTENSION TO ARBITRARY SILHOUETTE SHAPES

The second problem, mentioned in Section 2, is that the silhouette of the target object is limited to a simple rectangle. In this section we extend the former theory to deal with arbitrary silhouettes.

4.1 Decomposition of Silhouette

As the shape of the silhouette changes, value S_{unit}, the sum of SSA: $\hat{s}(x,y)$, diverges from rectangle case. In the first equation of equations(8),

$$S_{unit} = \iint_{F_1'} \frac{Tz^2}{(T_z-u)^2} dudv, \quad (17)$$

the range of integration F_1' can be a arbitrary silhouette. However, calculating this integral in an analytical way, as in the rectangle case, is not assured for all cases, and in many cases it is impractical. To allow arbitrary silhouettes to be handled in a practical way, let us consider a partial rectangle that floats on the floor, see Figure 6. The integration of equation(17) can be carried out in a way similar to that mentioned in Section2. Thus the value of the integration for the small rectangle j : ΔS^j_{unit} becomes

$$\Delta S^j_{unit} = d \cdot Tz^2 \left(\frac{1}{Tz-h_2} - \frac{1}{Tz-h_1}\right). \quad (18)$$

Assuming that the whole silhouette in the real world is divided into J partial rectangles, the S_{unit} value of the entire silhouette can be calculated as follows,

$$S_{unit} = \sum_{j=1}^{J} \Delta S^j_{unit} \quad (19)$$

Using equations(18) and (19) makes it practical to calculate the S_{unit} of an arbitrary silhouette.

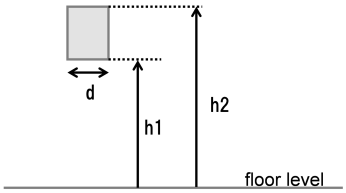

Figure 6: Partial rectangle.

4.2 Occlusion Model for Arbitrary Silhouette

The influence of occlusion strengthens with the number of people. When occlusion occurs, the foreground

pixels become smaller than when occlusion does not occur. This section explains how to add the influence of occlusion into the estimation algorithm mentioned above. The former paper(Arai et al., 2009) introduced an occlusion model for the simple rectangle case. In this paper the occlusion model is expanded to cover arbitrary silhouettes.

Figure 7 is used in the explanation of occlusion. Imagine triangle objects standing on the floor. To model the influence of occlusion, the density of objects ρ, i.e. number of objects per unit square on the floor plane, is considered. When pixel (x,y) is detected as foreground, it indicates that "at least one and ρ $\Omega(x,y)$ persons are present"; here $\Omega(x,y)$ is the area that corresponds to the set of possible positions, see reversed shadow region in Figure 7. Let $R(x,y)$ denote the redundancy of pixel(x,y) as follows,

$$R(x,y) = 1 + \rho\Omega(x,y). \tag{20}$$

From assumptions (b) and (c) given in 2.1, $\Omega(x,y)$ can be approximated as

$$\Omega(x,y) = S_0 \tan\theta, \tag{21}$$

here S_0 is the whole surface area of the silhouette in the real world, and θ is the angle of pixel(x,y), see Figure 7. If density ρ is given, and from the assumption of (d): positions of objects on the floor are random, the expectation value of observed surface area for one object, $\bar{S}$, becomes

$$\bar{S} \sim \frac{S_0}{\bar{R}} = \frac{S_0}{1 + \rho\bar{q}S_0}. \tag{22}$$

$\bar{R}$ is the expectation values of R(x,y), and $\bar{q}$ is the average value of tanθ in the region of interest(ROI). When the ROI on the floor is given beforehand and its area is A, the number of people can be expressed using ρ as follows,

$$N = \rho A. \tag{23}$$

We note that the observed surface area and expectation value of surface area for one person $\bar{S}$ have the following relationship,

$$\frac{\sum_{(x,y)\in F} \hat{s}(x,y)}{S_{unit}} = \frac{\bar{S}}{S_0} N. \tag{24}$$

From equations (22)(23)(24),

$$N \sim \frac{\nu}{(1 - \frac{\nu\bar{q}S_0}{A})}$$
$$\nu = \frac{\sum_{(x,y)\in F} \hat{s}(x,y)}{S_{unit}} \tag{25}$$

is derived.

The use of equation (25) permits number estimation with consideration of the influence of occlusion.

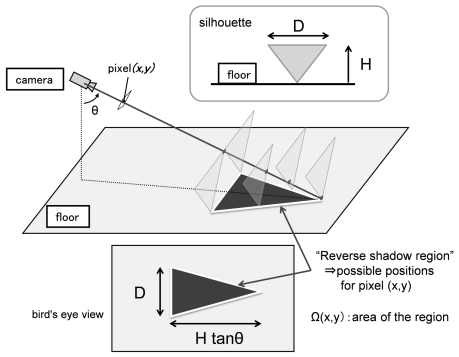

Figure 7: Geometric model of occlusion.

5 EXPERIMENTS AND RESULTS

To confirm the validity of the proposed method, two experiments were carried out. The first was to evaluate the validity of the slant rectangle model. The second was to confirm the method of silhouette decomposition. Both experiments used CG images generated by our simulation program.

5.1 Evaluation of Slant Rectangle Model

To confirm the validity and effectiveness of the slant rectangle model, an experiment was carried out using artificially generated binary images. To compare the slant model against the upright model, SSA was calculated for each model. Figure 8 shows examples of SSA calculated according to the upright rectangle model, and Figure 9 covers the slant rectangle model. The SSA of the upright rectangle model takes huge values at areas below the camera. On the other hand, the SSAs of the slant rectangle model do not significantly differ.

Cuboid images were generated to test the validity of the proposed method. In this experiment, only one cuboid is generated, and the distance from camera was changed incrementally changed(see Figure 10). Figure 11 shows the result of object number estimation. The horizontal axis corresponds to the lateral offset of the camera to the object, and vertical axis indicates the estimated number. The thick-dotted line indicates the result generated by the SSA value of the upright rectangle model as applied to the cuboid images. The thick-solid line indicates the result generated by the SSA value of the slant rectangle model.

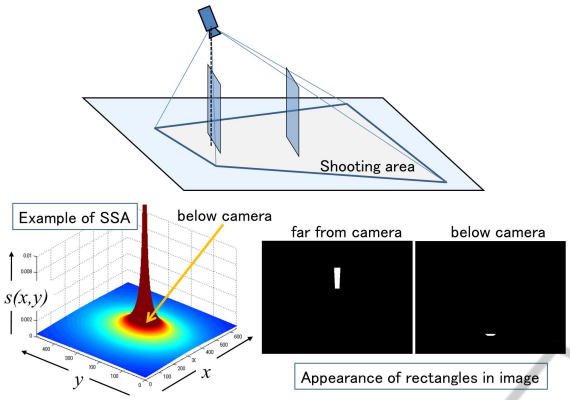

Figure 8: Example of SSA and CG-image (upright rect.).

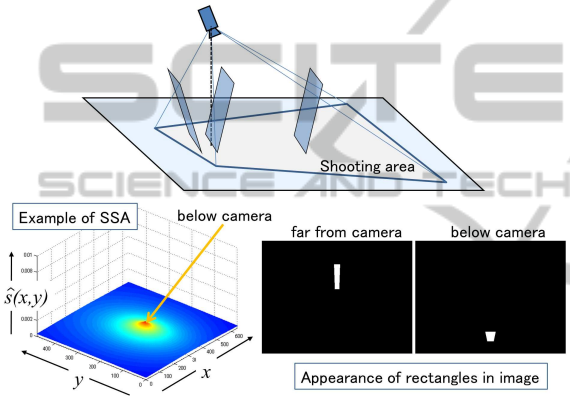

Figure 9: Example of SSA and CG-image (slant rect.).

Figure 10: Examples of cuboid images.

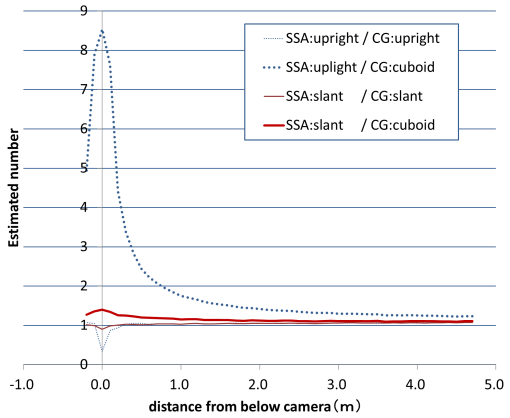

Figure 11: Estimation result (1-object case).

As can be seen, the upright model changes rapidly as the object approaches the camera. The slant model, on the other hand, changes little. This shows that the slant rectangle model can be applied to the directly downward case.

5.2 Evaluation of Silhouette Decomposition

To confirm the validity of the silhouette decomposition method, we conducted several experiments. Three types of silhouette images were generated, see Figure 12.

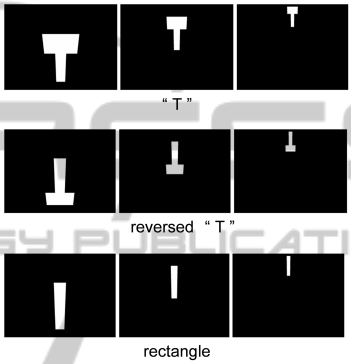

Figure 12: Examples of generated silhouettes.

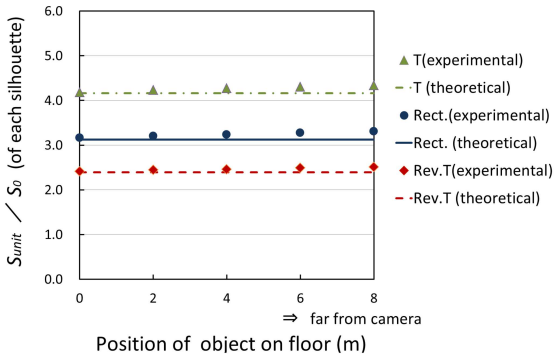

Figure 13: Invariance of sum of SSA: comparison of both theory and image processing.

Figure 13 shows the results of this experiment. The horizontal axis indicates the positions of the target objects, and vertical axis indicates S_{unit}/S_0 as given by the theoretical calculation, and $\sum_{(x,y)\in F} \hat{s}(x,y)/S_0$ by summing SSA values along with foreground pixels in generated image. As can be seen in Figure 13, each derived value is close to its theoretical value. The invariance of summed SSA, in the case of arbitrary silhouette, was confirmed by this experiment.

The second experiment was carried out to confirm

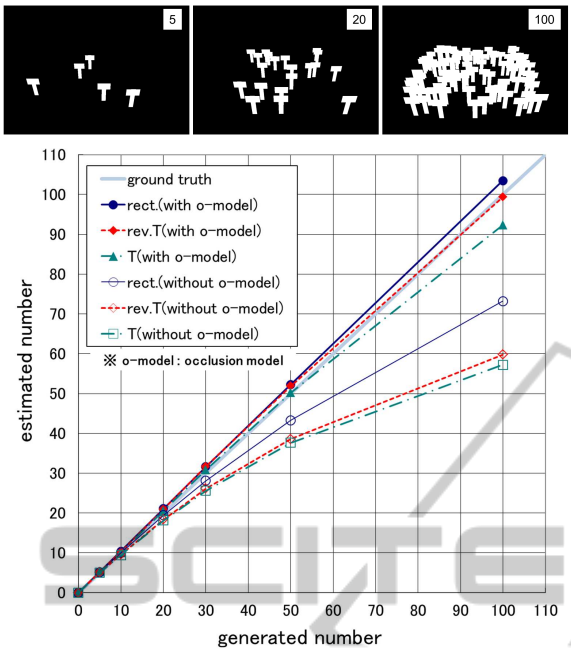

Figure 14: Estimation result in crowded situations.

the effectiveness of the proposed number estimation algorithm for arbitrary silhouettes. Three types of silhouette shapes with various numbers of silhouettes were generated, see examples in Figure 14. The horizontal axis plots the actual number of silhouettes generated, and the vertical axis indicates the estimated numbers. As can be seen, when the number of silhouettes is below 20 or 30, the estimated numbers without occlusion model are very close to the generated number, the ground-truth. On the other hand, with more than 30, the lines of estimated numbers without occlusion model deviate from the ground-truth. Applying the occlusion model mentioned in 4.2 makes the estimated numbers approach the ground-truth.

6 CONCLUSIONS

In this paper, we extended the basic theory of surface area analysis by introducing two improved techniques to better estimate the number of objects in images. We extended the basic theory to be able to deal with the directly downward capture case and with arbitrary silhouettes. The slant silhouette analysis realizes reliable estimation even if the object is directly under a camera. The silhouette decomposition technique extends object shape from a simple rectangle to arbitrary silhouettes. Experiments showed the validity of the proposed theory and the effectiveness of our crowd estimation technique.

REFERENCES

Antonini, G. and Thiran, J. (2006). Counting pedestrians in video sequences using trajectory clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(issue 8):1008–1020.

Arai, H., Miyagawa, I., Koike, H., and Haseyama, M. (2009). Estimating number of people using calibrated monocular camera based on geometrical analysis of surface area. *IEICE Transactions*, 92-A(8):1932–1938.

Cho, S.-Y., Chow, T., and Leung, C.-T. (1999). A neural-based crowd estimation by hybrid global learning algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, 29(PartB)(issue 4):535–541.

Kong, D., Gray, D., and Tao, H. (2006). A viewpoint invariant approach for crowd counting. *Proceedings of the 18th International Conference on Pattern Recognition*, 03:1187–1190.

Marana, A., Costa, L., Lotufo, R., and Velastin, S. (1998). On the efficacy of texture analysis for crowd monitoring. *Proceedings of the International Symposium on Computer Graphics, Image Processing*, 6:3521–3524.

Min, L., Zhaoxiang, Z., Kaiqi, H., and Tieniu, T. (2008). Estimating the number of people in crowded scenes by mid based foreground segmentation and head-shoulder detection. *International Conference on Pattern Recognition(ICPR 2008)*, pages 1–4.

Rabaud, V. and Belongie, S. (2006). Counting crowded moving objects. *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:705–711.

Sheng-Fuu Lin, J.-Y. C. and Chao, H.-X. (2001). Estimation of number of people in crowded scenes using perspective transformation. *IEEE Transactions on Man and Cybernetics*, 31(PartA)(Issue 6):645–654.

Sidla, O., Lypetskyy, Y., Brandle, N., and Seer, S. (2006). Pedestrian detection and tracking for counting applications in crowded situations. *Proceedings of the IEEE International Conference on Video and Signal Based Surveillance(AVSS06)*, pages 70–.

Wen, Q., JIA, C., Yu, Y., Chen, G., Yu, Z., and Zhou, C. (2011). People number estimation in the crowded scenes using texture analysis based on gabor filter. *Journal of Computational Information Systems*, 7:11:3754–3763.

Zhao, T., Nevatia, R., and Wu, B. (2007). Segmentation and tracking of multiple humans in crowded environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1198–1211.