skills of a professional calligrapher, it can do creative jobs such as making new works of art. Further, the robot can instruct people in the study of calligraphy, thus help to preserve and develop this culture. This is the final goal we are working towards.

References

1. Y. Adachi, S. Nakanishi, Y. Kuno, N. Shimada and Y. Shirai: Intelligent wheelchair using visual information from human face, J. of the Robotics Society of Japan, vol. 17, no. 3, pp. 423-431, 1999.
2. G. Bourhis and P. Pino: Mobile robotic and mobility assistance for people with motor impairments: Rational justification for the VAHM project, IEEE Trans. Rehab. Eng., vol.4, no. 1, pp. 7-12, 1996.
3. J. Hasegawa, H. Koshimizu, A. Nakayama and S. Yokoi, ed.: Image Processing on Personal Computer, Gijyutsu-Hyoron Co. Ltd., Tokyo, 1986.
4. M. Hirose, T. Takenaka, H. Gomi and N. Ozawa: Humanoid robot, J. of the Robotics Society of Japan, vol. 15, no. 7, pp. 23-25.
5. H. Ishiguro, T. Ono, M. Imai, T. Maeda, T. Kanda and R. Nakatsu: Robovie: A robot generates episode chains in our daily life, Proc. of 32nd Int. Symp. on Robotics, Korea, April, 2001, vol. 2, pp. 1356-1361.
6. R. L. Madarasz, L. C. Heiny, R. F. Cromp, and N. M. Mazur: The design of an autonomous vehicle for the disabled, *IEEE J. Robotics and Automat.* vol. RA-2, no. 3, 1986.
7. N. Matsuda, ed.: Gotaijikan, Kashiwashobo Publishing Co. Ltd., 1996.
8. M. Mazo, F.J. Rodriguez, J. Lazaro, J. Urena, J.C. Garcia, E. Santiso, P. Revenga and J.J.Garcia.: Wheelchair for physically disabled people with voice, ultrasonic and infrared sensor control, *Autonomous Robot*, vol. 2, pp. 203-224, 1995.
9. E. Prassler, J. Scholz, and P. Piorini: Navigating a robotic wheelchair in a railway station during rush hour, *Int. J. Robotics Res.*, vol. 18, no. 7, pp. 711-727, Jul. 1999.
10. T. Sakai, R. Hirata, S. Okada, N. Hiraki, Y. Okajima, N. Tanaka and S. Uchida.: Rehabilitation robot for stroke patients (TEM, therapeutic exercise machine), Proc. of 32nd Int. Symp. on Robotics, Korea, 19-21 April 2001, vol. 3, pp. 1587-1591.
11. A. Takanishi: Human communication oriented humanoid robot, J. of the Robotics Society of Japan, vol. 15, no. 7, pp. 11-14.
12. F.H. Yao, G.F. Shao, H. Yamada and K. Kato: The Visual Feedback System for an Outdoor Intelligent Powered Wheelchair, Proc. of 32nd Int. Symp. on Robotics, Korea, vol. 1, pp. 592-597, April 2001.
13. http://www.honda.co.jp/ASIMO
14. http://www.sony.co.jp/SonyInfo
15. http://www.aibo.com/
16. http://www.robocup.org/
17. http://www.sok.co.jp/r_d/index.html

Automated People Counting using Template Matching and Head Search

Grantham Kwok-Hung Pang, Chi-Kin Ng,
Department of Electrical and Electronic Engineering,
The University of Hong Kong.

Abstract

People counting using image processing has been carried out for years. Conventional methods can count people accurately when only a few isolated people pass through a counting region in a non-crowded situation. In this paper, the emphasis is on people counting in a crowded environment and a method using head search and model matching is described. A camera is mounted vertically downwards viewing the people heads from the top. People head search can be used to locate some passengers. In addition, templates obtained from the perspective projection of the human model are used to locate and isolate individual person. Our approach aims at dealing with a congested situation where occlusion is a major problem. This paper describes a real-time, high-accuracy, automated people counting system that has been developed. Experimental results are illustrated and the effectiveness of the developed method for real-time application is verified.

Keywords: *Automated people counting, real-time image processing, template matching.*

1. INTRODUCTION

Everyday, a large number of people move around in all directions in buildings, on roads, railway platforms and stations. The information on passenger flow is very important to public transport operators and it can help them to control the flow and manage the traffic effectively.

One conventional people counting method is based on turnstiles, but the mechanical contact of turnstiles is inconvenient and uncomfortable to the passengers. It is also impossible to install the gates on every platform and escalator to count the number of people. Counting method using optical beam [1] fails to count correctly when several passengers cross the beam at the same time.

People counting based on image processing on either the spatial images [2-7] or spatial-temporal images [8-9] have been proposed. For spatial images, the counting methods included mathematical morphology [2], shape model filtering on the round shape of people head [3], block matching with clustering [4], supervised split and merge [5] and optical flow [6,7]. Mathematical morphology with averaging-thresholding methods [8] and template matching in stereo images [9] are

used to count the number of people in spatial-temporal images. Their counting accuracy is satisfactory when passenger flow rate is low. When people walk in a crowded group, the accuracy can be greatly decreased.

In this paper, an automated people counting system that can run in real-time on a PC-based platform is described. The algorithm is based on a novel two-stage people isolation method using head search and projected model templates. The system can provide high counting accuracy even in a crowded and occluded situation.

2. SYSTEM OVERVIEW

The people counting system is targeted to operate in an indoor environment with limited variation of the illumination level. A color digital camera is placed vertically above the passengers viewing downwards to the head. The camera is mounted near the ceiling, which is at around 3 to 5 meters from the floor. The detection region viewed from the camera would be larger than 2m along a walking direction. Normal plain floor background is assumed. The aberration and spherical distortion of the camera is assumed negligible. The frame rate has been set to 5 fps, but it can vary according to the hardware configuration. People would walk in all directions with a nominal speed of around 1.5m/s. In the developed algorithm, a person should be seen for five or more successive frames in an image sequence. After image acquisition, the system would consist of the following stages: image segmentation, cluster isolation, cluster tracking and counting.

3. IMAGE SEGMENTATION

Segmentation is used to discriminate the people from the background. The input for segmentation are images of M pixels in column and N pixels in row. Each pixel is a 24-bit true-color RGB pixel. In order to reduce the computation in the latter stage, the segmented image is processed in sub-blocks consisting of BxB pixels.

Before segmentation, threshold values are found as follows. Two background images are acquired from the scene. They are converted into the HSV color model, and the maximum range of variation of the hue (H) value is obtained. Hue is used for the representation of the color feature of the background for segmentation because it is invariant to images containing high saturation, even in the presence of shading, shadows and highlights [10].

Hence, in order to segment an image, the image is converted into the HSV color space. When the hue value of a pixel within each sub-block of the image lies outside the threshold value range of $[H_{min},H_{max}]$, the corresponding pixel is considered as part of a person. Otherwise, the pixel is considered as background or shadow. When more than 50% of the pixels within a sub-block is occupied, the sub-block will be considered as part of a person. Otherwise, the sub-block is considered as background or shadow.

4. CLUSTER ISOLATION

From a segmented image, the number of people and their positions should be isolated to perform counting. In a crowded situation, the isolation of an individual person is a difficult problem. The occlusion due to perspective projection will

make the isolation even more difficult. A two-stage method based on head search and template matching has been developed. Combining the two methods, cluster isolation is divided into head search, template model matching and cluster removal as shown in Fig. 1.

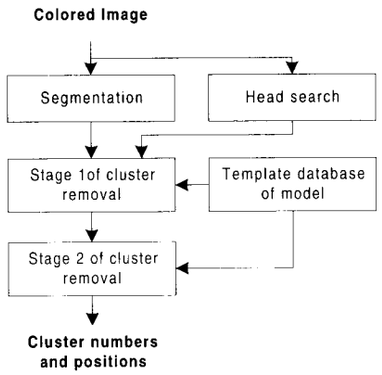

Fig. 1. Block diagram of cluster isolation

4.1 Head search

Head search is used to find the potential locations of people heads. As the camera is viewing vertically downwards, the heads are usually seen circular in shape. The hair color can help in the search (e.g. dark or black color in most part of Asia; silvery grey, brown or golden color in North America). However, it is only a heuristic approach that is used in stage one of cluster removal. If this heuristic method fails, we still have stage 2 of cluster removal to identify people. In our test scene, the people hair is mostly dark in color. The head search is thus based on finding the circular regions that are dark.

The color image is first divided into sub-blocks, where the sub-block mean is calculated by summing up each RGB color value within each sub-block. The sub-blocked mean image will be convolved with a circular template mask. The resulting data represents the sum of the color values over the circular template mask. When the sums are below a threshold value (T_{head}), the regions are considered as heads and the regional centres will be treated as the head centres.

4.2 Template database

The usual template matching technique with standard masks of the template can isolate people, but it fails to isolate the varying size of people due to the perspective projection and occlusion. Therefore, in this paper, template masks from the human model is proposed. Simple geometric shapes are used to model people, and their perspective projections on the view plane are obtained as templates. These templates can be used to find and isolate each cluster from the segmented image where each cluster represents a person.

In order to obtain the perspective projection, a person is modeled by a set of simple geometric shapes. In this paper, a person is modeled by an ellipsoid as head,

used to count the number of people in spatial-temporal images. Their counting accuracy is satisfactory when passenger flow rate is low. When people walk in a crowded group, the accuracy can be greatly decreased.

In this paper, an automated people counting system that can run in real-time on a PC-based platform is described. The algorithm is based on a novel two-stage people isolation method using head search and projected model templates. The system can provide high counting accuracy even in a crowded and occluded situation.

2. SYSTEM OVERVIEW

The people counting system is targeted to operate in an indoor environment with limited variation of the illumination level. A color digital camera is placed vertically above the passengers viewing downwards to the head. The camera is mounted near the ceiling, which is at around 3 to 5 meters from the floor. The detection region viewed from the camera would be larger than 2m along a walking direction. Normal plain floor background is assumed. The aberration and spherical distortion of the camera is assumed negligible. The frame rate has been set to 5 fps, but it can vary according to the hardware configuration. People would walk in all directions with a nominal speed of around 1.5m/s. In the developed algorithm, a person should be seen for five or more successive frames in an image sequence. After image acquisition, the system would consist of the following stages: image segmentation, cluster isolation, cluster tracking and counting.

3. IMAGE SEGMENTATION

Segmentation is used to discriminate the people from the background. The input for segmentation are images of M pixels in column and N pixels in row. Each pixel is a 24-bit true-color RGB pixel. In order to reduce the computation in the latter stage, the segmented image is processed in sub-blocks consisting of BxB pixels.

Before segmentation, threshold values are found as follows. Two background images are acquired from the scene. They are converted into the HSV color model, and the maximum range of variation of the hue (H) value is obtained. Hue is used for the representation of the color feature of the background for segmentation because it is invariant to images containing high saturation, even in the presence of shading, shadows and highlights [10].

Hence, in order to segment an image, the image is converted into the HSV color space. When the hue value of a pixel within each sub-block of the image lies outside the threshold value range of $[H_{min},H_{max}]$, the corresponding pixel is considered as part of a person. Otherwise, the pixel is considered as background or shadow. When more than 50% of the pixels within a sub-block is occupied, the sub-block will be considered as part of a person. Otherwise, the sub-block is considered as background or shadow.

4. CLUSTER ISOLATION

From a segmented image, the number of people and their positions should be isolated to perform counting. In a crowded situation, the isolation of an individual person is a difficult problem. The occlusion due to perspective projection will

make the isolation even more difficult. A two-stage method based on head search and template matching has been developed. Combining the two methods, cluster isolation is divided into head search, template model matching and cluster removal as shown in Fig. 1.

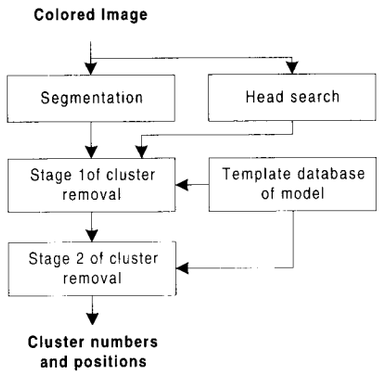

Fig. 1. Block diagram of cluster isolation

4.1 Head search

Head search is used to find the potential locations of people heads. As the camera is viewing vertically downwards, the heads are usually seen circular in shape. The hair color can help in the search (e.g. dark or black color in most part of Asia; silvery grey, brown or golden color in North America). However, it is only a heuristic approach that is used in stage one of cluster removal. If this heuristic method fails, we still have stage 2 of cluster removal to identify people. In our test scene, the people hair is mostly dark in color. The head search is thus based on finding the circular regions that are dark.

The color image is first divided into sub-blocks, where the sub-block mean is calculated by summing up each RGB color value within each sub-block. The sub-blocked mean image will be convolved with a circular template mask. The resulting data represents the sum of the color values over the circular template mask. When the sums are below a threshold value (T_{head}), the regions are considered as heads and the regional centres will be treated as the head centres.

4.2 Template database

The usual template matching technique with standard masks of the template can isolate people, but it fails to isolate the varying size of people due to the perspective projection and occlusion. Therefore, in this paper, template masks from the human model is proposed. Simple geometric shapes are used to model people, and their perspective projections on the view plane are obtained as templates. These templates can be used to find and isolate each cluster from the segmented image where each cluster represents a person.

In order to obtain the perspective projection, a person is modeled by a set of simple geometric shapes. In this paper, a person is modeled by an ellipsoid as head,

a cylindroid as the body and a flat ellipse on the top of the cylindroid as the shoulder (Fig. 2).

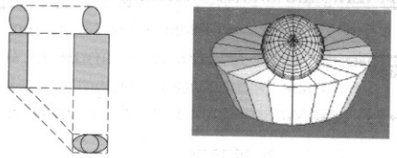

Fig. 2. Two side views and top views of a model of the passenger (left) and the three-dimensional view of the model (right)

To find the perspective projection of the model into the view plane, the camera and the scene are modeled as shown in Fig. 3.

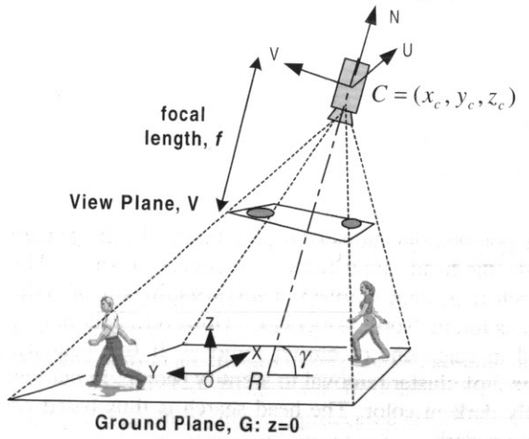

Fig. 3. Modeling of camera and the scene for finding perspective projection of object points

The image viewed from the camera is divided into (P-1) columns and (Q-1) rows to form a grid. There are, in total, P×Q grid points. Their corresponding positions on the 3D space are calculated. The world co-ordinate system in XYZ is transformed into camera co-ordinate system in UVN. It can be done by translating XYZ world space into the view point C and then rotating to match the UVN view space.

$$\begin{pmatrix} u \\ v \\ n \\ 1 \end{pmatrix} = \mathbf{M} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\gamma & \sin\gamma & 0 \\ 0 & -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -x_c \\ 0 & 1 & 0 & -y_c \\ 0 & 0 & 1 & -z_c \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

View point C is a distance of f in front of the plane of projection along the N-axis. Suppose the view rectangle at the view plane is bounded from $-V_u$ to V_u in the U-axis and from $-V_v$ to V_v in the V-axis. Grid points can be represented as (p, q) where $p \in [1..P]$ and $q \in [1..Q]$. Each grid point corresponding to a point (u_p, v_q) on the view plane can be calculated as follows:

$$(u_p, v_q) = ((\frac{p-1}{P-1} - \frac{1}{2}) * V_u, (\frac{q-1}{Q-1} - \frac{1}{2}) * V_v)$$

At each position on the P×Q grid points, the model projects from the 3D space into the view plane in a perspective way. Head position and body position in the view plane are found as the projected position of the ellipsoid centre of the head model, and the bottom center of the cylindroid of the body model respectively. The templates with their corresponding head and body positions are stored in a template database.

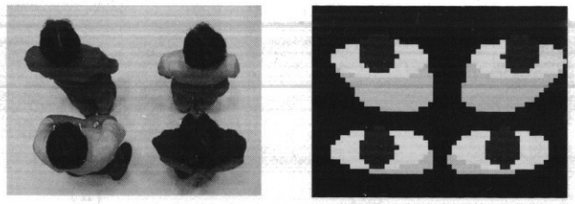

Fig. 4. Cluster isolation using human model templates

4.3 Cluster removal

Cluster removal is performed in two stages. In the first stage, the template with the head position nearest to the positions from the head search process are used to remove the occupied regions from the segmented image. In the second stage, the remaining clusters are searched, located and removed by using the templates in the template database. An example of people and their corresponding human model templates are shown in Fig. 4.

In the first stage, the template with the nearest distance between the head position of the template and the potential head position in the head search process is used. The nearest distance is calculated by the following formula:

$$D = \min_{i \in [1..P], j \in [1..Q]} \sqrt{(xh_{i,j} - xh)^2 + (yh_{i,j} - yh)^2}$$

where (xh, yh) is the head position in segmented image and $(xh_{i,j}, yh_{i,j})$ is the head position of the template which is the nearest to the head position.

From the head position of the template, the corresponding body position of the template can be retrieved from the template database. The sub-blocks of the segmented image covered by the template are removed. The steps would be continued until all the templates corresponding to the potential positions of the heads are used.

a cylindroid as the body and a flat ellipse on the top of the cylindroid as the shoulder (Fig. 2).

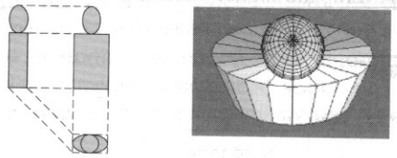

Fig. 2. Two side views and top views of a model of the passenger (left) and the three-dimensional view of the model (right)

To find the perspective projection of the model into the view plane, the camera and the scene are modeled as shown in Fig. 3.

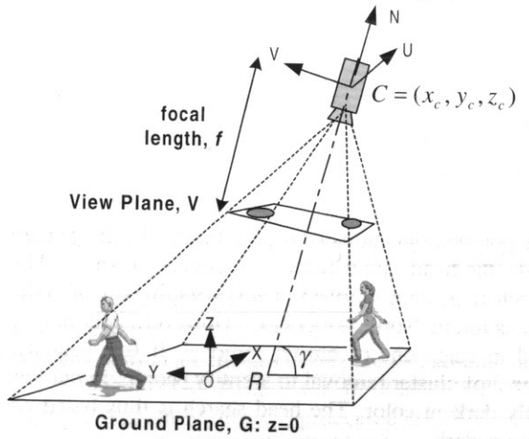

Fig. 3. Modeling of camera and the scene for finding perspective projection of object points

The image viewed from the camera is divided into (P-1) columns and (Q-1) rows to form a grid. There are, in total, P×Q grid points. Their corresponding positions on the 3D space are calculated. The world co-ordinate system in XYZ is transformed into camera co-ordinate system in UVN. It can be done by translating XYZ world space into the view point C and then rotating to match the UVN view space.

$$\begin{pmatrix} u \\ v \\ n \\ 1 \end{pmatrix} = \mathbf{M} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\gamma & \sin\gamma & 0 \\ 0 & -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -x_c \\ 0 & 1 & 0 & -y_c \\ 0 & 0 & 1 & -z_c \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

View point C is a distance of f in front of the plane of projection along the N-axis. Suppose the view rectangle at the view plane is bounded from $-V_u$ to V_u in the U-axis and from $-V_v$ to V_v in the V-axis. Grid points can be represented as (p, q) where $p \in [1..P]$ and $q \in [1..Q]$. Each grid point corresponding to a point (u_p, v_q) on the view plane can be calculated as follows:

$$(u_p, v_q) = ((\frac{p-1}{P-1} - \frac{1}{2}) * V_u, (\frac{q-1}{Q-1} - \frac{1}{2}) * V_v)$$

At each position on the P×Q grid points, the model projects from the 3D space into the view plane in a perspective way. Head position and body position in the view plane are found as the projected position of the ellipsoid centre of the head model, and the bottom center of the cylindroid of the body model respectively. The templates with their corresponding head and body positions are stored in a template database.

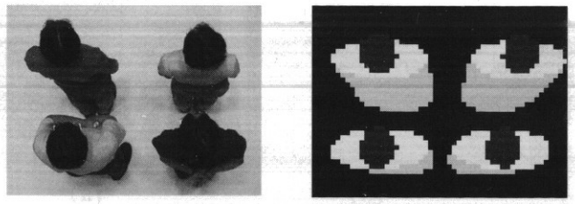

Fig. 4. Cluster isolation using human model templates

4.3 Cluster removal

Cluster removal is performed in two stages. In the first stage, the template with the head position nearest to the positions from the head search process are used to remove the occupied regions from the segmented image. In the second stage, the remaining clusters are searched, located and removed by using the templates in the template database. An example of people and their corresponding human model templates are shown in Fig. 4.

In the first stage, the template with the nearest distance between the head position of the template and the potential head position in the head search process is used. The nearest distance is calculated by the following formula:

$$D = \min_{i \in [1..P], j \in [1..Q]} \sqrt{(xh_{i,j} - xh)^2 + (yh_{i,j} - yh)^2}$$

where (xh, yh) is the head position in segmented image and $(xh_{i,j}, yh_{i,j})$ is the head position of the template which is the nearest to the head position.

From the head position of the template, the corresponding body position of the template can be retrieved from the template database. The sub-blocks of the segmented image covered by the template are removed. The steps would be continued until all the templates corresponding to the potential positions of the heads are used.

In the second stage, the sub-block values of the template at (p, q) position is multiplied with the corresponding sub-block values on the remained binary image and the sum is obtained (Fig. 5a). Totally, there are P×Q sums when all the templates are used (Fig. 5b). These sums indicate the probability of presence of a person within the template mask. In order to reduce the effect of occlusion due to the body viewed from the camera, the template has higher weighting on the head and shoulder rather than the body, so the template values are 3 for head, 2 for shoulder, 1 for body and 0 for all others. If the sum is larger than a certain coverage percentage (CP), it should be considered as the presence of a person.

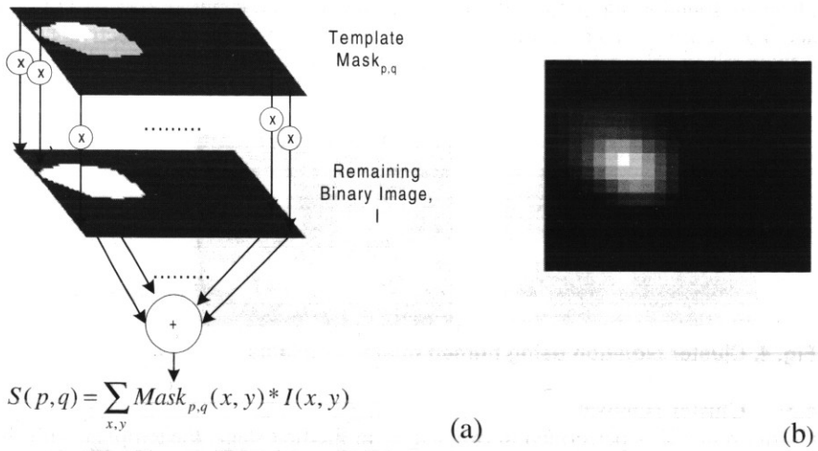

Fig. 5(a) The sum calculation between template at (p, q) and remained binary image; (b) Sums of all S(p, q)

5. CLUSTER TRACKING AND COUNTING

The tracking algorithm is based on a matching process of the tokens between two successive frames. It includes estimation and matching. A token is also determined to be a coming-in token, coming-out token, tracking token or false detection. People counting is done by counting the number of tokens coming out of the tracking region. A tracking of the tokens between two successive time instants is carried out. Estimation of the token positions is by linear prediction from previous walking displacement. Matching is performed between the tokens within the tracking and alerting regions. People are counted as coming out of the tracking region when they pass through the top or down count line. The determination of a new comer is by a coming-in token in the tracking region. False detection is alarmed when a token in the previous time instant no longer matches any token in the current time instant within the tracking region. The test results show that the algorithm is capable of tracking the trajectory of the people and thus count the number of people.

6. EVALUATION AND RESULTS

The output of the people counting system is the counting results, which give the number of people passed through a counting region within a specified period of time, and the direction of travel. The developed system has been evaluated in many different scenarios, and in both crowded and non-crowded cases, with satisfactory results. The result in one scenario is given here. It is also found that the method can be implemented in real-time on a PC-733 MHz platform.

An image sequence obtained at five frames per second is processed by the system. The image resolution is 320x240. The snapshots of the scenario are shown in Fig. 6. The type of background is a corridor and the people can travel bi-directionally. The camera is mounted 5m from the floor and the coverage area is 4.14m(W)x3.09m(H).

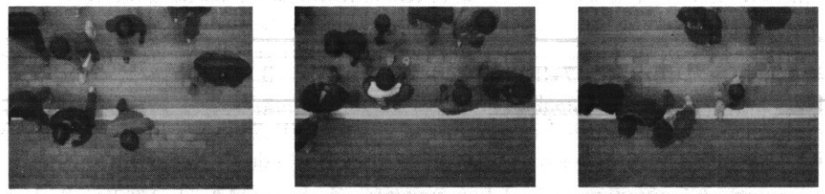

Fig. 6. Snapshots of one test scenario

Some test results are shown below:

Condition	Number of frames	Expected no. of counting	No. of counting from the system	Absolute error	Absolute error with direction of travel	Overall counting error
Crowded Connected Uni-directional walking	1-894	Top=0 Down=234 Total=234	Top=6 Down=213 Total=219 False detection =47	Top=+6 Down=-21 Total=-15	11.53%	6.41%
Crowed Connected Bi-directional walking	1-459	Top=54 Down=59 Total=113	Top=51 Down=61 Total=112 False detection =17	Top=-3 Down=2 Total=-1	4.42%	0.88%
				Average	7.96%	3.65%

Table 1. Test results in a scenario with original background

Most of the errors are mainly due to the poor segmentation of the people. Also, when many people come together, their shadows sometimes have significant changes to the background. At the moment, the tracking algorithm uses simple geometric information to estimate and match tokens between two frames. A missing token along the path of walk would generate false detection and cause under-counting.

In the second stage, the sub-block values of the template at (p, q) position is multiplied with the corresponding sub-block values on the remained binary image and the sum is obtained (Fig. 5a). Totally, there are P×Q sums when all the templates are used (Fig. 5b). These sums indicate the probability of presence of a person within the template mask. In order to reduce the effect of occlusion due to the body viewed from the camera, the template has higher weighting on the head and shoulder rather than the body, so the template values are 3 for head, 2 for shoulder, 1 for body and 0 for all others. If the sum is larger than a certain coverage percentage (CP), it should be considered as the presence of a person.

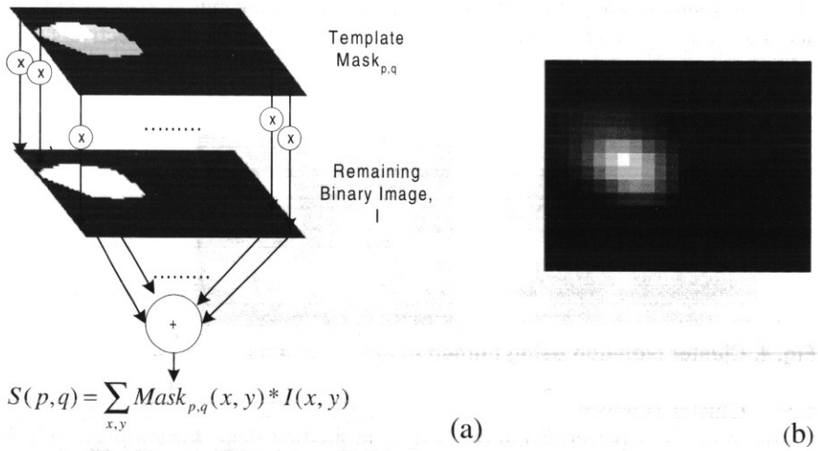

Fig. 5(a) The sum calculation between template at (p, q) and remained binary image; (b) Sums of all S(p, q)

5. CLUSTER TRACKING AND COUNTING

The tracking algorithm is based on a matching process of the tokens between two successive frames. It includes estimation and matching. A token is also determined to be a coming-in token, coming-out token, tracking token or false detection. People counting is done by counting the number of tokens coming out of the tracking region. A tracking of the tokens between two successive time instants is carried out. Estimation of the token positions is by linear prediction from previous walking displacement. Matching is performed between the tokens within the tracking and alerting regions. People are counted as coming out of the tracking region when they pass through the top or down count line. The determination of a new comer is by a coming-in token in the tracking region. False detection is alarmed when a token in the previous time instant no longer matches any token in the current time instant within the tracking region. The test results show that the algorithm is capable of tracking the trajectory of the people and thus count the number of people.

6. EVALUATION AND RESULTS

The output of the people counting system is the counting results, which give the number of people passed through a counting region within a specified period of time, and the direction of travel. The developed system has been evaluated in many different scenarios, and in both crowded and non-crowded cases, with satisfactory results. The result in one scenario is given here. It is also found that the method can be implemented in real-time on a PC-733 MHz platform.

An image sequence obtained at five frames per second is processed by the system. The image resolution is 320x240. The snapshots of the scenario are shown in Fig. 6. The type of background is a corridor and the people can travel bi-directionally. The camera is mounted 5m from the floor and the coverage area is 4.14m(W)x3.09m(H).

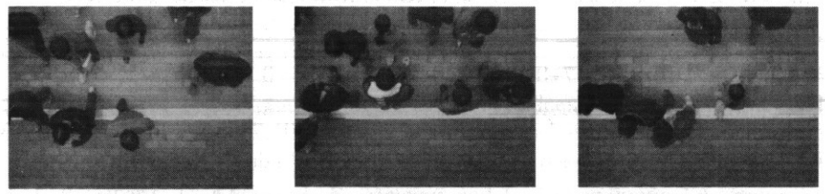

Fig. 6. Snapshots of one test scenario

Some test results are shown below:

Condition	Number of frames	Expected no. of counting	No. of counting from the system	Absolute error	Absolute error with direction of travel	Overall counting error
Crowded Connected Uni-directional walking	1-894	Top=0 Down=234 Total=234	Top=6 Down=213 Total=219 False detection =47	Top=+6 Down=-21 Total=-15	11.53%	6.41%
Crowed Connected Bi-directional walking	1-459	Top=54 Down=59 Total=113	Top=51 Down=61 Total=112 False detection =17	Top=-3 Down=2 Total=-1	4.42%	0.88%
				Average	7.96%	3.65%

Table 1. Test results in a scenario with original background

Most of the errors are mainly due to the poor segmentation of the people. Also, when many people come together, their shadows sometimes have significant changes to the background. At the moment, the tracking algorithm uses simple geometric information to estimate and match tokens between two frames. A missing token along the path of walk would generate false detection and cause under-counting.

In a more extensive system evaluation, the average counting error is 4.23%. This result is comparable to or better than other similar approaches in a spatial domain. Especially in crowed situation with people walking bi-directionally, the above is a high-accuracy result. For comparison, the accuracies achieved by other researchers in [2], [3], [4] are around 94%, 87% and 90% respectively.

The developed algorithm has also been analysed for real-time performance. The worst-case computational time for all the system modules is given below:

1. Image acquisition t_{acq} : $2.3x10^{-6}$s
2. Segmentation t_{seg} : 0.05s
3. Head search t_{head} : 0.00637s
4. Cluster removal by human model templates (one person) $t_{cluster}$: 0.00263s
5. Tracking and counting (for 40 tokens) $t_{trackcount}$: $85.7x10^{-6}$ s

Assuming that the maximum number of people in each image is 40 (N_{max}). The time required in the worst case for processing one image in the sequence is

$$= t_{acq} + t_{seg} + t_{head} + N_{\max} * t_{cluster} + t_{trackcount}$$

$$=(2.3\times10^{-6}+ 0.05+0.00637+0.00263*40+85.7\times10^{-6})s = 0.1617s$$

Hence, the system is capable of processing over six frames per second. The performance was evaluated using an Intel Pentium III 733MHz PC with 256MB RAM, running Visual C++ 6.0 under the Microsoft Windows 98 platform.

7. CONCLUSIONS

An automated people counting system has been developed. The algorithm of the system is mainly based on a novel two-stage people isolation to find the positions of individual person across an image sequence. The first stage of isolation would perform head search to locate some people using the features of circular black regions. The second stage of isolation uses model templates to locate the remaining people. These model templates are obtained from the perspective projection of a three-dimensional human model into a two-dimensional image plane. The use of the projected model templates considers the shape variation of the people located at different positions of a scenario.

The developed method has been tested with images captured at real scenarios at 5m camera-mounting height relative to a plain floor. People can walk in a bi-directional way with different degree of crowdedness. The overall system counting accuracy is around 95%. The worst-case computation time of the algorithm is also evaluated for the real-time feasibility of the system running on PC.

References

1. L. Khoudour, L. Duvieubourg and J. P. Deparis, "Real-time passenger counting by active linear cameras", Proceedings of SPIE on real time imaging, v. 2661, 1996, pp. 106-117.

2. R. Glachet, S. Bouzar, F. Lenoir and J. Blosseville, "Counting Pedestrian in the Subway Corridors using Image Processing", Proceedings of SPIE, Applications of Digital Image Processing XVIII, v. 2564, 1994, pp. 261-270.

3. X. Zhang and G. Sexton, "Automatic human head location for pedestrian counting", The Sixth Int. Conference on Image Processing and its Applications, vol. 2, 1997, pp. 535 –540.

4. M. Rossi and A. Bozzoli, "Tracking and counting moving people", Proceedings of IEEE International Conference on Image Processing 1994 (ICIP-94), vol. 3, 1994, pp. 212 –216.

5. A. Rourke and M. G. H. bell, "An image-processing system for pedestrian data collection", IEEE Conf. on Road Traffic Monitoring and Control, April 1994, pp. 123-126.

6. F. Bartolini, V. Cappellini and A. Mecocci, "Counting People Getting In and Out of a Bus by Real-time Image-sequence Processing", Image and Vision Computing, Vol. 12, No. 1, 1994, pp. 36-41.

7. P. Nesi and A. Del Bimbo, "A vision system for estimating people flow", Jorge L. C. Sanz, "Image Technology", Berlin, Springer-Verlag, 1996, pp. 170-201.

8. J. P Deparis, L. Khoudour, B. Meunier and L. Duvieubourg, "A device for counting passengers making use of two active linear cameras: comparison of algorithms", Proceedings of IEEE Int. Conf. on Systems, Man and Cyb., vol. 3, 1996, pp. 1629 –1634.

9. K. Terada, D. Yoshida, S. Oe and J. Yamaguchi, "A Method of Counting the Passing People by Using the Stereo Images", IEEE Conf. on Image Processing, 1999, pp. 338- 342.

10. F. Perez and C. Koch, "Toward Color Image in Analog VLSI: Algorithm and Hardware", International Journal of Computer Vision, Vol. 12, 1994, pp17-42.

In a more extensive system evaluation, the average counting error is 4.23%. This result is comparable to or better than other similar approaches in a spatial domain. Especially in crowed situation with people walking bi-directionally, the above is a high-accuracy result. For comparison, the accuracies achieved by other researchers in [2], [3], [4] are around 94%, 87% and 90% respectively.

The developed algorithm has also been analysed for real-time performance. The worst-case computational time for all the system modules is given below:

1. Image acquisition t_{acq} : $2.3x10^{-6}$s
2. Segmentation t_{seg} : 0.05s
3. Head search t_{head} : 0.00637s
4. Cluster removal by human model templates (one person) $t_{cluster}$: 0.00263s
5. Tracking and counting (for 40 tokens) $t_{trackcount}$: $85.7x10^{-6}$ s

Assuming that the maximum number of people in each image is 40 (N_{max}). The time required in the worst case for processing one image in the sequence is

$$= t_{acq} + t_{seg} + t_{head} + N_{\max} * t_{cluster} + t_{trackcount}$$

$$=(2.3\times10^{-6}+ 0.05+0.00637+0.00263*40+85.7\times10^{-6})s = 0.1617s$$

Hence, the system is capable of processing over six frames per second. The performance was evaluated using an Intel Pentium III 733MHz PC with 256MB RAM, running Visual C++ 6.0 under the Microsoft Windows 98 platform.

7. CONCLUSIONS

An automated people counting system has been developed. The algorithm of the system is mainly based on a novel two-stage people isolation to find the positions of individual person across an image sequence. The first stage of isolation would perform head search to locate some people using the features of circular black regions. The second stage of isolation uses model templates to locate the remaining people. These model templates are obtained from the perspective projection of a three-dimensional human model into a two-dimensional image plane. The use of the projected model templates considers the shape variation of the people located at different positions of a scenario.

The developed method has been tested with images captured at real scenarios at 5m camera-mounting height relative to a plain floor. People can walk in a bi-directional way with different degree of crowdedness. The overall system counting accuracy is around 95%. The worst-case computation time of the algorithm is also evaluated for the real-time feasibility of the system running on PC.

References

1. L. Khoudour, L. Duvieubourg and J. P. Deparis, "Real-time passenger counting by active linear cameras", Proceedings of SPIE on real time imaging, v. 2661, 1996, pp. 106-117.

2. R. Glachet, S. Bouzar, F. Lenoir and J. Blosseville, "Counting Pedestrian in the Subway Corridors using Image Processing", Proceedings of SPIE, Applications of Digital Image Processing XVIII, v. 2564, 1994, pp. 261-270.

3. X. Zhang and G. Sexton, "Automatic human head location for pedestrian counting", The Sixth Int. Conference on Image Processing and its Applications, vol. 2, 1997, pp. 535 –540.

4. M. Rossi and A. Bozzoli, "Tracking and counting moving people", Proceedings of IEEE International Conference on Image Processing 1994 (ICIP-94), vol. 3, 1994, pp. 212 –216.

5. A. Rourke and M. G. H. bell, "An image-processing system for pedestrian data collection", IEEE Conf. on Road Traffic Monitoring and Control, April 1994, pp. 123-126.

6. F. Bartolini, V. Cappellini and A. Mecocci, "Counting People Getting In and Out of a Bus by Real-time Image-sequence Processing", Image and Vision Computing, Vol. 12, No. 1, 1994, pp. 36-41.

7. P. Nesi and A. Del Bimbo, "A vision system for estimating people flow", Jorge L. C. Sanz, "Image Technology", Berlin, Springer-Verlag, 1996, pp. 170-201.

8. J. P Deparis, L. Khoudour, B. Meunier and L. Duvieubourg, "A device for counting passengers making use of two active linear cameras: comparison of algorithms", Proceedings of IEEE Int. Conf. on Systems, Man and Cyb., vol. 3, 1996, pp. 1629 –1634.

9. K. Terada, D. Yoshida, S. Oe and J. Yamaguchi, "A Method of Counting the Passing People by Using the Stereo Images", IEEE Conf. on Image Processing, 1999, pp. 338- 342.

10. F. Perez and C. Koch, "Toward Color Image in Analog VLSI: Algorithm and Hardware", International Journal of Computer Vision, Vol. 12, 1994, pp17-42.