

Loughborough University Institutional Repository

Automated people-counting by using low-resolution infrared and visual cameras

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: AMIN, I.J. ... et al, 2008. Automated people-counting by using low-resolution infrared and visual cameras. *Measurement*, 41 (6), pp. 589 - 599.

Additional Information:

- This article was published in the journal, *Measurement* [© Elsevier]. The definitive version is available at: <http://dx.doi.org/10.1016/j.measurement.2007.02.010>

Metadata Record: <https://dspace.lboro.ac.uk/2134/15693>

Version: Submitted for publication

Publisher: © Elsevier Ltd.

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the published version.

Automated people counting by using low-resolution infrared and visual cameras

I. J. Amin, A. J. Taylor, F. Junejo, A. Al-Habaibeh, , R. M. Parkin

The Wolfson School of Mechanical and Manufacturing Engineering

Loughborough University

Loughborough

Leicestershire

LE11 3TU

Abstract

Non-contact counting of people in a specified area has many applications for safety, security and commercial purposes. Visible sensors have inherent limitations for this task, being sensitive to variations in ambient lighting and colours in the scene. Infrared imaging can overcome many of these problems but normally hardware costs are prohibitively expensive. A system for counting people in a scene using a combination of low cost, low resolution visual and infrared cameras is presented in this paper. The aim of this research was to assess the potential accuracy and robustness of systems using low resolution images. This approach results in considerable savings on hardware costs, enabling the development of systems which may be implemented in a wide range of applications. The results of eighteen experiments show that the system can be accurate to within 3% over a wide range of lighting conditions.

Keywords: people-counting, imaging, infrared, artificial neural network (ANN)

Notation

p	Energy Radiated
λ	Wavelength
T	Temperature (Kelvin)
h	Plank's Constant
c_c	Velocity of light
b	Boltzman Constant
w	Radiated energy
ε	Emissivity
η	Boltzman constant
$P(q)$	Interpolated value
$L_i(q)$	Lagrange polynomial
q	Point at which interpolation takes place
f_i	Known values on the grid at points (q_i)
t_i	Desired or target response for ith unit
y_i	Actually produced response for ith unit
E	Error for a single training pattern in neural network
Ω	Transfer function output value
o_{th}	Output thresholded image
α_{th}	Thresholding value
r	Original image
d	Constant
c	Constant
q	Grayscale image
x	Infrared image
δ	Average body heat
m	Thresholded infrared image

1. Introduction

Automated counting is an active research topic in many areas including biology [1], medicine, quality control and industrial machine vision processes amongst others. There are many situations where it is useful or essential to count people and numerous automated people-counting systems have been developed over the years. A variety of contact based sensors are in use, such as pedestrian barriers on entrances to public buildings and gateways. Most commercially available non-contact based counters use infrared beams or ultrasonic sensors, and specialized human information sensors are also developed for this task [2]. However the most commonly used non-contact system still remains the visual camera [3-5].

One present disadvantage with visual counting systems is the cost – a high spatial resolution visual camera and a frame grabber required for the system are still fairly expensive items. However a more fundamental problem, even with high spatial resolution cameras, remains the inaccuracy associated with visual detection of people. If a person is wearing the same shades of grey as the background it is difficult to distinguish between the background and the clothes. Also there are no reliable ways of distinguishing with accuracy a person from similar objects. These objects in the background are one of the main concerns, commonly raising false alarms in many automated people counting systems. Generally it can be said that background separation is not an easy task. Furthermore visual automated counting systems can only work in the presence of ambient lighting such as in an office environment, sunlight, or other types of lighting. In case of emergencies, such as fire or blackouts, the system will malfunction during evacuation of the building and thus could be rendered useless at crucial times. Similar situations can occur with exterior use of visual people counters [5], there will be false alarms during night time if there is no special lighting arrangement in the area under consideration.

Thus a system is proposed to overcome these problems by using a low cost infrared thermal imager together with a visual camera. The visual camera uses an image-processing algorithm that can distinguish between people and objects with an accuracy of about 12%.

This system, developed by Schofield et al [4], uses visual automated counting but can be modified easily to accommodate low spatial resolution visual images. The working principle is based on the background training of visual images using a neural network.

2. Thermal Imaging

Thermal infrared (IR) imaging sensors respond to emitted, more than reflected, radiation. All objects emit heat by three means - conduction, convection and radiation. Conduction transfers heat through solid objects, convection transfers heat through fluids and radiation transfers heat through electromagnetic radiation.

Objects continuously radiate heat with certain wavelengths, dependent upon the temperature of the radiating object and its spectral emissivity. As the object temperature increases the radiation increases. The radiation emitted includes the infrared emission which consists of electromagnetic wavelengths between $0.7\text{ }\mu\text{m}$ and $100\text{ }\mu\text{m}$. Small ranges of infrared emission from the objects are detected by the thermal imager and then made visible as an image in the form of a thermogram – a mapping of apparent temperatures.

The concept behind infrared emission detection of the thermal imager is the assumption that a black body is a perfect radiator; it emits and absorbs all incident energy. The energy emission for the black body is the greatest possible for energy emission for that particular temperature. Radiation power emitted by a black body as given by Plank's radiation law [6] is:

$$p(\lambda, T) = \frac{2\pi hc^2}{\lambda^2} \left[\exp\left(\frac{hc}{\lambda bT}\right) - 1 \right]^{-1} \quad [1]$$

Real objects are not perfect emitters or absorbers. Thus emissivity (ε) of the real surface is defined as the ratio of thermal radiation emitted by a surface at given temperature to that of a black body for the same temperature, spectral and directional conditions [7, 8]. Thus the emissivity of a black body is 1 and all other real surface emissivities will be between 1 and 0.

According to the Stefan Boltzman Law of emissivity radiation:

$$w = \varepsilon \eta T^4 \quad [2]$$

Thermal imaging converts thermal radiation into a digital signal which is then converted into a visible image. This study uses a newly developed thermal imager of type IRYSIS IRI 1001. This offers many advantages including low cost, a wide temperature measurement range and the capability to capture images on an IBM-PC via an RS-232C port. The thermal imager is housed in an aluminium casing of 100 mm by 100 mm complete with optics, pyroelectric detector [9], chopping motor and optical modulator. It has a temperature measurement range of -20 to 90°C with an accuracy of $\pm 0.1^\circ\text{C}$ [10]. Although it is a low resolution, 16×16 pixel, thermal imager it can be used to display images of up to 128×128 pixels using bilinear or bicubic interpolation. The interpolation process estimates values of intermediate components of continuous function in discrete samples. An interpolation technique does not add extra information into the image but can provide better thermal images for human perception. For bicubic interpolation, the output pixel value is the weighted average of the pixels in the nearest 4×4 neighbourhood. Mathematically, bicubic interpolation can be described as follows:

Let L_i be a third degree polynomial. The Lagrange polynomial interpolation is given by [11]

$$P(q) = \sum_{i=0}^3 f_i L_i(q) \quad [3]$$

Previous research by Schofield et al [4] has shown that low resolution visual images give similar visual information to that of high resolution devices, as the visual information will be processed by computer. A similar approach is used for low-resolution thermal images. A low resolution thermal imager will cost much less than a typical high resolution thermal imager, around one tenth of the cost, and will be much smaller than a conventional thermal imager. Additionally the low resolution imager is specially designed for embedded systems, where

data can be directly streamed through an RS232 connection to the computer for on-line monitoring and off-line analysis.

3. Neural Networks

An artificial neural network (ANN) is an information-processing paradigm inspired by the way in which the densely interconnected, parallel structure of the human brain processes information. Neural networks resemble the human brain in the following two ways:

A neural network acquires knowledge through learning.

A neural network's knowledge is stored within inter-neuron connection strengths known as synaptic weights.

Artificial neural networks are collections of mathematical models that emulate some of the observed properties of biological nervous systems and draw on the analogies of adaptive biological learning. The key element of the ANN paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements that are analogous to *neurons* and are tied together with weighted connections that are analogous to *synapses*.

The main advantage of using a neural network is the full automation of the learning and classification processes, allowing them to be implemented in fully automated monitoring systems, such as people counting, to recognize and classify different patterns without human involvement.

Neural networks are composed of simple elements operating in parallel. As in nature, the network function is determined largely by the connections between elements. Some Neural networks are classified as feed-forward while others are recurrent (i.e., implement feedback) depending on how data is processed through the network. Another way of classifying neural network types is by their method of learning or training, as some employ *supervised training* while others are referred to as *unsupervised* or *self-organizing* networks. The selection of a supervised or unsupervised network is largely dependent on the data to be processed for the

training of the network. More detailed information is available in a number of sources [12, 13].

During supervised learning of an ANN an input stimulus is applied that results in an output response. Then this response is compared with a desired output i.e. the target response. If the actual response differs from the target response, the neural network generates an error signal. A popular measure of the error 'E' for a single training pattern, is the sum of square differences i.e. [14].

$$E = \frac{1}{2} \sum_i (t_i - y_i)^2 \quad [4]$$

The error "E" is then used to calculate the adjustment that should be made to the network's synaptic weights so that the actual output matches the target output.

In contrast to supervised learning the case of unsupervised learning does not require a teacher; i.e. no target output is required. It is usually found in the context of recurrent and competitive nets. In the case of unsupervised learning there is no separation of the training set into input and output pairs during the training session, the neural net receives as its input many different excitations, or input patterns, and it arbitrarily organizes the patterns into categories. When a stimulus is later applied the neural net provides an output response indicating the class to which the stimulus belongs. If a class cannot be found for the input stimulus, a new class is generated. However, it should be noted that even though unsupervised learning does not require a teacher, it requires *guidelines* to determine how it will form groups. Grouping may be based on shape, colour, or material consistency or on some other property of the object [14, 15].

In this study, a sequential (RAM based) neural network has been used which uses binary weights, i.e. 0/1 values, stored in RAM memory blocks which themselves play the role of the 'neurons' in the system. This approach, sometimes called '*weightless neural computing*', has many advantages over other neural networks, including fast network training. It uses 'one-

shot' learning procedures very different from the iterative ones of conventional neural networks and furthermore they can operate well on low resolution images [13, 16].

4. Experimental work

The experiment was conducted by mounting the low-cost visual imaging device (Webcam) and the IRISYS IRI1001 thermal imager looking vertically down. Markers were placed on the floor under consideration so that both infrared imager and visual camera were sharing the same information. The visual imager has a much wider field of view than the thermal imager, thus only a cropped visual view was taken into consideration.

Three control experiment scenarios were used. Each scenario was based on six experiments with differences in position, movement of subjects and different lighting conditions. The background images with no subjects were also taken each time. Each experiment conducted contained around 150 visual and infrared samples of data stored on a hard disk. The length of each experiment varied from 3 to 5 minutes depending upon the subjects involved, and during all experiments the data acquisition software was kept running. The three scenarios were as follows:

4.1. Elevator camera (static)

In this scenario ten volunteers were involved which resulted in thirty tests. This simulates the elevator surveillance camera with a restriction of any volunteer leaving the scene during the length of each test. During each test volunteers were asked to stand for five seconds at random positions in the area being monitored. Also the number of volunteers increased as the test progressed. The maximum number of volunteers in tests was five and each test was repeated five times with a random selection of volunteers.

4.2. Gate Camera

This scenario simulates the gate camera for counting. Volunteers were asked to enter the scene from one side and leave on the opposite side. Thirty tests were conducted, with each

test repeated five times with a random selection of volunteers. Two special conditions, i.e. one person standing within the gate for a certain period of time and one person stopping and returning to where he/she entered from, were included in these tests.

4.3. Elevator camera (dynamic)

This scenario simulates the actual elevator surveillance camera. The volunteers in this scenario were allowed to leave and enter the scene but only from one side which is the elevator door. The maximum number of volunteers in the tests was ten. During a test volunteers were given specific instructions when to enter or leave the scene. It also simulates peak times as well as off-peak times during the day. As in the previous scenarios all volunteers were selected randomly for each test to maintain the validity of the final result.

5. Image processing strategy

The visual system used is lower cost than traditional CCTV cameras, around 1/10th of the cost. The low cost CMOS sensor used by the visual system also develops a noise factor, which presents a major issue to be considered during the visual analysis. Thus images with simple subtraction with respect to the reference scene do not provide a consistent image in our case which can be thresholded.

The visual analysis carried out is very similar to that done by Schofield et al [4] except that the equipment used is low cost. The thermal imaging analysis is also done separately. The results of each analysis are then further compared to increase the accuracy of the system.

These showed that the system can be developed to be capable of counting in a smoked filled room and other emergency situations, which is not the case with conventional visual counting systems. In the following sections both visual and infrared data are analysed separately and then the combined results are discussed.

5.1. Visual Analysis

The visual analysis uses the background identification technique employed by Schofield et al [4]. This system avoids any standard approach which fails to take light variations into account, hence it is independent of light intensities in the image. Thus this process is chosen for visual analysis for the development of our system with some modifications, for example we do not require location information in an image as it is not necessary in the proposed application. The visual counting system developed should have the following characteristics:

Accuracy	Approximately 10%
Error	Maximum of +/-1 error in 4 to 10 people in a scene
Lighting conditions	Adaptable to any indoor lighting conditions
Adaptability	Most scenes in indoor buildings

Table 1. Design guidelines for the visual counting system

The accuracy is measured by the comparison of simulation data and the training set for that particular experiment.

5.1.1. Stage 1a: Pre-Processing

The pre-processing stage for visual analysis consists of resizing and thresholding. The initial image acquired from the experiment is 288x288 pixels. This is then reduced to 72x72 pixels. The reduction in resolution allows faster processing and a faster counting rate with negligible degradation in the thresholding result. For example, for the initial image of 288x288 pixels thresholding takes about 4.5 seconds using a fast processing speed while 72x72 pixels takes only about 2.5 seconds using MatLAB. This will improve significantly after final development of the system using a programming language such as C or C++.

Following resizing a reference image from each experiment is taken. Reference images are merely background images with no people in the scene. These reference images are thresholded not by the constant greyscale value but by applying adaptive local thresholding.

The neighbouring pixel will allow the intensity of pixels to be compared with each other. If the comparison of these pixels is high, up to a certain value set by another variable ' α_{th} ', the pixel is turned black otherwise white. Thus it can be mathematically expressed as:

$$[o_{th}]_{(i,j)} = \begin{pmatrix} 1 \text{ if } \alpha_{th} > |r_{(i,j)} - r_{(i-1,j-1)}| \\ 0 \text{ if } \alpha_{th} \leq |r_{(i,j)} - r_{(i-1,j-1)}| \end{pmatrix} \quad [5]$$

Here ' α_{th} ' is the global thresholding value of the image being processed. To calculate this value $1/3^{rd}$ pixel values of images are randomly selected. The difference of intensities of these pixels is taken from their diagonal neighbour. Here two constants 'c' and 'd' are introduced in the thresholding value of ' α_{th} '. After summing all of the intensity difference values the final value is multiplied by a constant 'c', which is less than 1. The value acquired is then added to the constant value of 'd'. The thresholding expression is mathematically expressed as:

$$\alpha_{th} = d + c \sum |q_{(k,l)} - q_{(k-1,l-1)}| \quad [6]$$

The optimal values of 'c' and 'd' are found by experimenting with the visual images taken during the experiment.

5.1.2. Stage 2a: Background Identification

The background identification is based on the RAM based neural network creation and training of that network. Only background images are trained using this network.

The thresholded image is divided into 4x4-sections, with 18 sections in each row and 324 altogether in the 72x72 pixel image. We consider each 4x4-section containing 16 pixels divided further into four sections, which are termed sub-sections. These sub-sections, containing four pixels each, are then randomly selected, and this selection remains the same over the life of a neural network. These randomly selected sub-sections are used as the

addresses of RAM. For each 4x4 section created and randomly selected 4 sub-sections create a single classifier.

Training of a RAM based network is done by reading the 4 pixels from each group in a 4x4 section then outputting 1 to the RAM of that certain address as shown in Figure 5. For example if the value of the 4 randomly selected pixels is 0101 then it outputs 1 to the corresponding memory output of that address. Then it starts summing up all values in the memory addresses, which are specific for each individual 4 pixel group. Thus for every section of the image seen it outputs 1 into the RAM of that section address. It goes on until all the background samples are trained for that network. There is no reason to run the samples again through the network for a background already seen, as the result will always be the same for that particular image.

To simulate the image using a trained network a thresholded sample of the image is fed into the network. The sample image is then divided into the random sections, which are the same as that of the trained network. The addresses of sample images are compared with the trained network values. If the network has already seen the same section during training it outputs '1', if the network hasn't seen anything like the section it outputs '0'. An output image is constructed with 1's as the background and 0's as the unseen object during the training. After inverting the image the unseen objects or people then appear as a cluster of 1's in that image. 51 reference samples are used for training of the RAM-based neural network.

A 5x5 section is scanned over the output image by the neural network. For highest counts found in 5x5 sections in the image a count is incremented, and the 3x3 section in the middle is set to zero whereas the 16 outside values are halved. This process is continued until a certain cut-off value is achieved for the image. An optimized cut-off value is found by comparing the result found with the actual result.

5.2. Infrared Analysis

For the development of the low-resolution infrared counting system certain guidelines were laid down as follows:

The infrared analysis system developed will be used in conjunction with the visual system but can be used as a stand-alone system with very slight modifications.

Accuracy	Approximately 5%
Error	Maximum of +/-1 error in 4 to 10 people in a scene
Adaptability	To most indoor building conditions and objects in scene (except extreme temperature conditions, e.g. +50°Celsius)
Lighting Variations	Completely insensitive to lighting variations in a scene

Table 2. Design guidelines for the infrared counting system

5.2.1. Stage 1b: Pre-processing

Infrared data taken from the experiment are taken offline into MATLAB. The raw infrared data taken from the experiment is interpolated to find the ‘average body heat’. The temperature of a person is generally higher than the background, except in very hot areas such as desert, but as this experiment is conducted inside a building we can assume a reasonably consistent temperature difference. Average heat of the background image in this experiment is found to be:

$$\text{averageheat} = \frac{\sum \text{pixels}}{256} \cong 24.3^{\circ} \text{Celsius}$$

The internal temperature of the IRISYS® infrared camera remains at 32.375 ° Celsius.

Thus the overall body temperature ranges for the duration of our experiment remain within:

Minimum *bodytemp* = 27 ° C

Maximum *bodytemp* = 32 ° C

Average δ = 29.5 ° C

The ‘average body heat’ calculated from the infrared data is then used as the thresholding value for the experiments conducted. This value varies with weather conditions and location of the experiment, such as whether it is conducted indoors or outdoors.

Infrared images of 16x16 pixels are processed using the following equation:

$$[m]_{(j,k)} = \begin{cases} 1 & \text{if } x \geq \delta \\ 0 & \text{if } x < \delta \end{cases}$$

x is an element of the original matrix of 16x16 elements from the infrared imager, m is an element of the thresholded matrix.

The infrared images after thresholding at average body heat give a distinguishable result that can be used for object recognition. But this is true only for small numbers of people as when the area under consideration becomes crowded then the algorithm becomes unreliable and hence further processing is necessary.

5.2.2. Stage 2b: Back Propagation Neural Network

For infrared image counting neural network areas are selected as the images are small, up to 16x16 pixels. After thresholding the infrared images are trained on back propagation neural networks.

The specification for the final network selected is as follows:

Inputs	256
Hidden Layer	1
Hidden Layer Neurons	280
Hidden Layer Function	Sigmoid Function
Output Layer Neurons	1
Output Layer Function	PureLin Function
Training Performance goal achieved	0.00642496
Epochs	500
Learning rate	0.005
Training Samples	360

Table 3. Configuration of optimized neural network for Infrared Analysis

Training of the backpropagation neural network is done by using twenty (20) samples from all eighteen (18) experiments as fed into the network.

6. Results

6.1.1. Infrared neural network simulation and results

The results acquired from the infrared data are plotted in the form of percentage error, with the error plot based on the simulation of 200 samples selected from each of 18 experiments. The error tends to increase as the number of people counted in the scene increases though there is much scatter and the error does not continue to rise as the numbers approach the maximum of ten.

6.1.2. Visual RAM based neural network simulation and result

For visual images the system accuracy is within 5% when fewer than six people are in each scene. However as the number of people in each scene increases the error percentage increases to around 12%. This is due to people standing very close to each other, as would be the case for example at peak time in elevators. To overcome these large errors in the system both infrared and visual results are combined.

6.1.3. Combined results of visual and infrared systems

It can be seen from the above results that an infrared system is capable of predicting a high density of people with high accuracy, whereas a visual system has proved to be more reliable for predicting lower densities of people. Therefore, in order to optimize the overall accuracy of the system, fusion of results from thermal and visual systems is carried out by taking percentage error and shifting the weight of results with less error percentage. As a result, as shown in Figure 9, the maximum percentage error has been reduced to 3%, even for scenarios containing a high density of people. A small reduction in density from ten to eight or less people in the scene reduces the error to within 1.5 %.

Discussion

The aim of this research was to assess the potential accuracy and robustness of people-counting systems using low resolution images. It has been shown that a combination of visual and infrared automated counting systems can be a cost-effective, practical and reliable means of counting people who are free to move within a defined area.

Combining the two automated counting methods, visual RAM based system and infrared system, has been shown to give significant improvements in accuracy compared with either system used singly. The maximum percentage error of 3% is far more accurate than either the visual system or infrared system alone, and remains within this value for more than eight

people in the experiment. The error of 1.5 % for lower numbers is considered to be sufficiently accurate for many applications.

Much of the value of this approach lies in its reliability and the capability to function in low light or completely dark conditions. Reliable non-contact counting of people will be essential to many future safety systems, and it is clearly crucial that counting systems will continue to work in emergency situations where power loss or thick smoke are likely to affect visibility. The low resolution infrared imager can provide slightly less accurate but very reliable counting capability in low or zero light conditions, making it ideally suited to emergency situations.

Future improvements are clearly possible with improved resolution and through developments in software and the application of artificial neural networks. In this research the visual image processing was carried out using MatLAB, and the thresholding time of 2.5 seconds can be much reduced in future by using a programming language such as C++. All imaging systems, whether visual or infrared, will encounter inherent difficulties in accurately counting higher numbers of people in close proximity. Low resolution infrared imaging offers major cost advantages and, although accuracy may be improved with higher resolution images, the higher resolution devices will be much more expensive for the foreseeable future. The accuracy achieved using only 16 by 16 pixels is considered to be sufficient for many practical applications. This research has assessed the accuracy and potential of the combined imaging approach and has demonstrated that the system works successfully in its current form.

References

1. J. Marotz, C. Lubbert, and W. Eisenbei, *Effective object recognition for automated counting of colonies in Petri dishes (automated colony counting)*. J. Computer Methods and Programs in Biomedicine, 2001. **66**((2-3)): p. 183-198.
2. K. Hashimoto, T.C. Kawaguchi, S. Matsueda, K. Morinaka, and N. Yoshiike, *People-counting system using multisensing application*. Sensors and Actuators A: Physical, 1998. **66**(1-3): p. 50-55.
3. T.W.S. Chow and S.Y. Cho, *Industrial neural vision system for underground railway station platform surveillance*. I Advanced Engineering Informatics, 2002. **16**(1): p. 73-83.
4. A.J. Schofield, T.J. Stonham, and P.A. Mehta, *Automated people counting to aid lift control*. Automation in Construction, 1997. **6**((5-6)): p. 437-445.
5. C. Sacchi, G. Gera, L. Marcenaro, and C.S. Regazzoni, *Advanced image-processing tools for counting people in tourist site-monitoring applications*. J. Signal Processing, 2001. **81**(5): p. 1017-1040.
6. S.G. Burnay, T.L. Williams, and C.H. Jones, *Applications of Thermal Imaging*. 1998, Bristol, Great Britain: I O P Publishing Ltd.
7. S. Kato, N. Minobe, and S. Tsugawa, *Applications of inter-vehicle communications to driver assistance system*. JSAE Review, 2003. **24**(1): p. 9-15.
8. G.C. Holst, *Common sense approach to thermal Imaging*. 2000: SPIE Optical Engineering Press.
9. J.L. Miller, *Principles of Infrared Technology: a practical guide to the state of the art*. 1994: Van Nostrand Reinhold.
10. *IRISYS: The Affordable Thermal Imager*, I.I.S.L.
11. A. Al-Habaibeh and R. M. Parkin, *An autonomous low-cost infrared system for the on-line monitoring of manufacturing processes using novelty detection*. International Journal of Advanced Manufacturing Technology, 2003. **22**((3-4)): p. 249-258.
12. J.A. Anderson, *An Introduction to Neural Networks*. 1995: The MIT Press.
13. S. Haykin, *Neural Networks: A Comprehensive Foundation, Second ed.* 1999: Prentice Hall.
14. C. Demant, B. Streicher-Abel, and P. Waszkewitz, *Backpropagation training, in Industrial Image Processing: Visual Quality Control in Manufacturing*. 1999: Springer-Verlag, Telos.
15. R.C. Gonzalez and R. E. Woods, *Object recognition, in Digital Image Processing*. 2002: Prentice Hall.
16. N. S. Martinelli and R. Seoane. *Automotive night vision system*. in *SPIE - The International Society for Optical Engineering, Proceedings of Thermosense XXI*. 1999.

List of figures

Figure 1. Mathematical expressions for the transfer functions.

Figure 2. Experimental setup and data acquisition system.

Figure 3. Overview of the people-counting system.

Figure 4. The scene with no people (upper two images) and the scene containing eight people (lower two images).

Figure 5. Example of RAM neural network training.

Figure 6. The backpropagation neural network.

Figure 7. Infrared image neural network error percentage.

Figure 8. Visual RAM based simulation result comparison.

Figure 9. Combined percentage error of visual and infrared systems.

Figure 1. Mathematical expressions for the transfer functions.

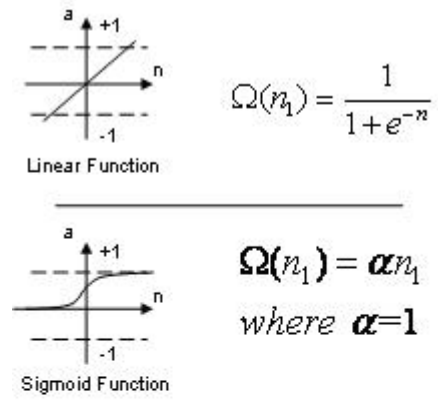


Figure 2. Experimental setup and data acquisition system.

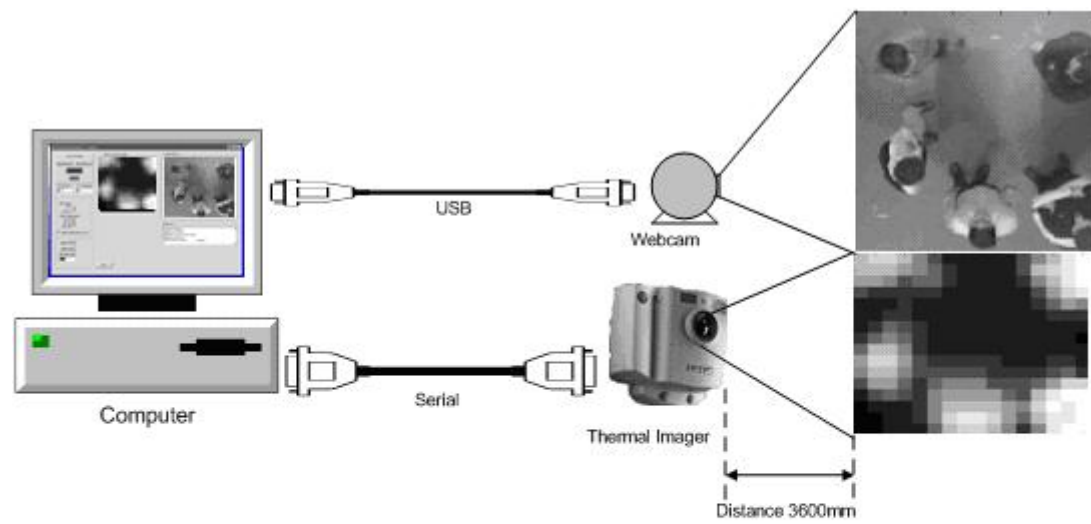


Figure 3. Overview of the people-counting system.

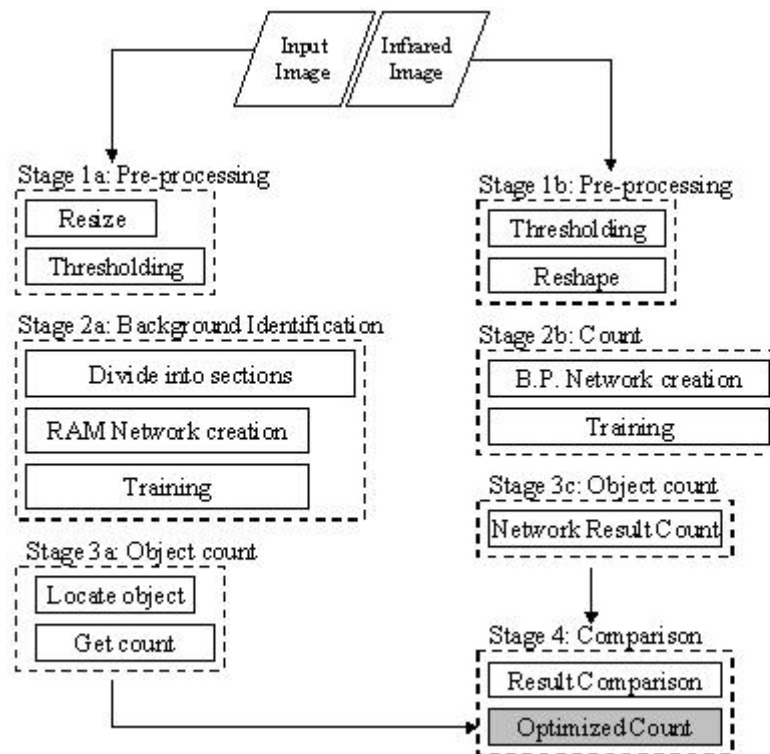


Figure 4. The scene with no people (upper two images) and the scene containing eight people (lower two images).

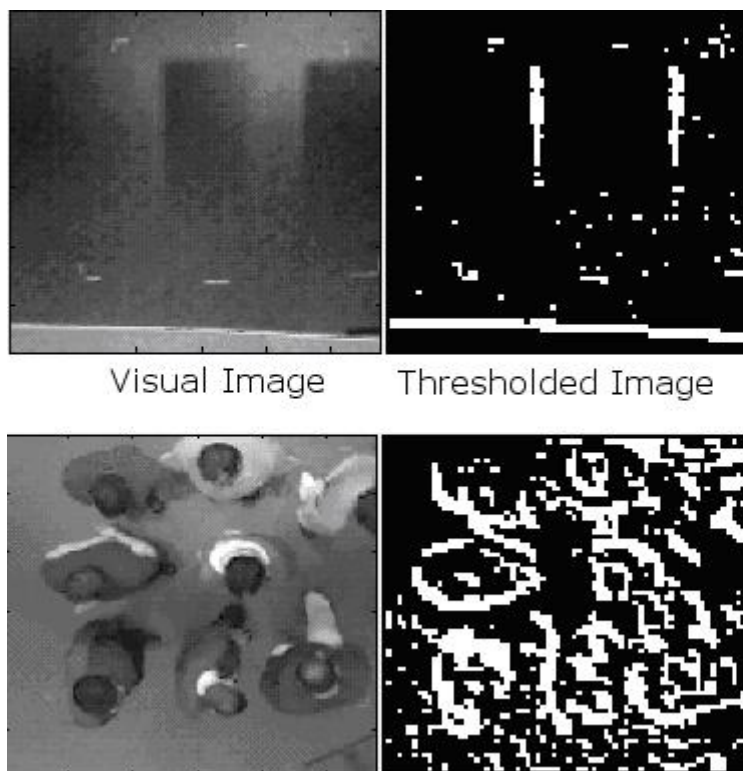


Figure 5. Example of RAM neural network training.

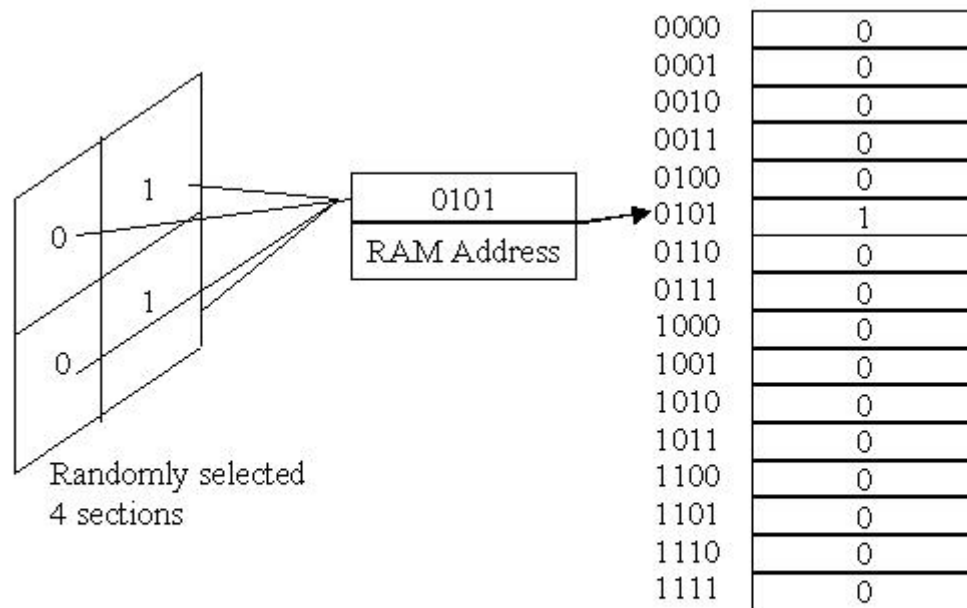


Figure 6. The backpropagation neural network.

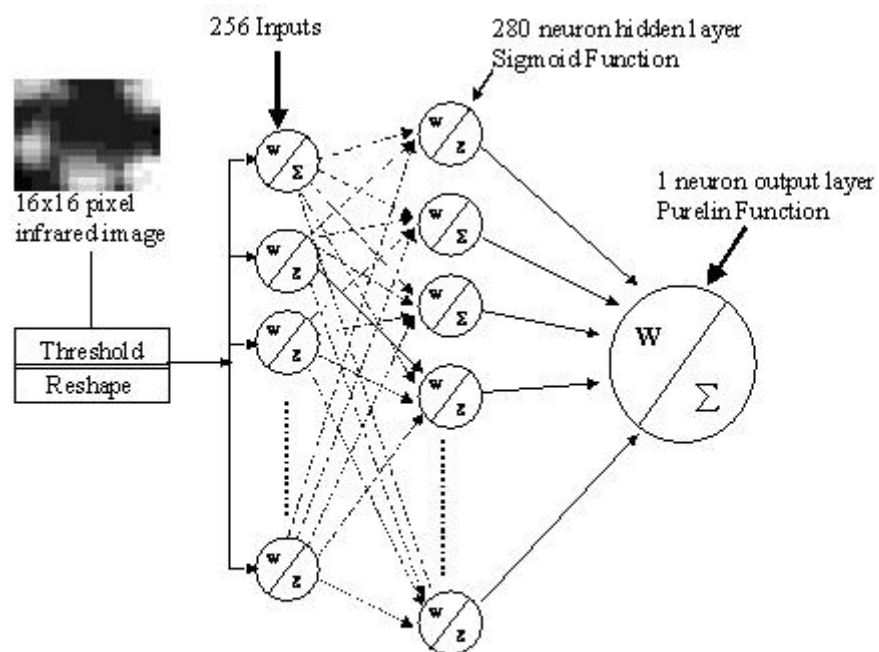


Figure 7. Infrared image neural network error percentage.

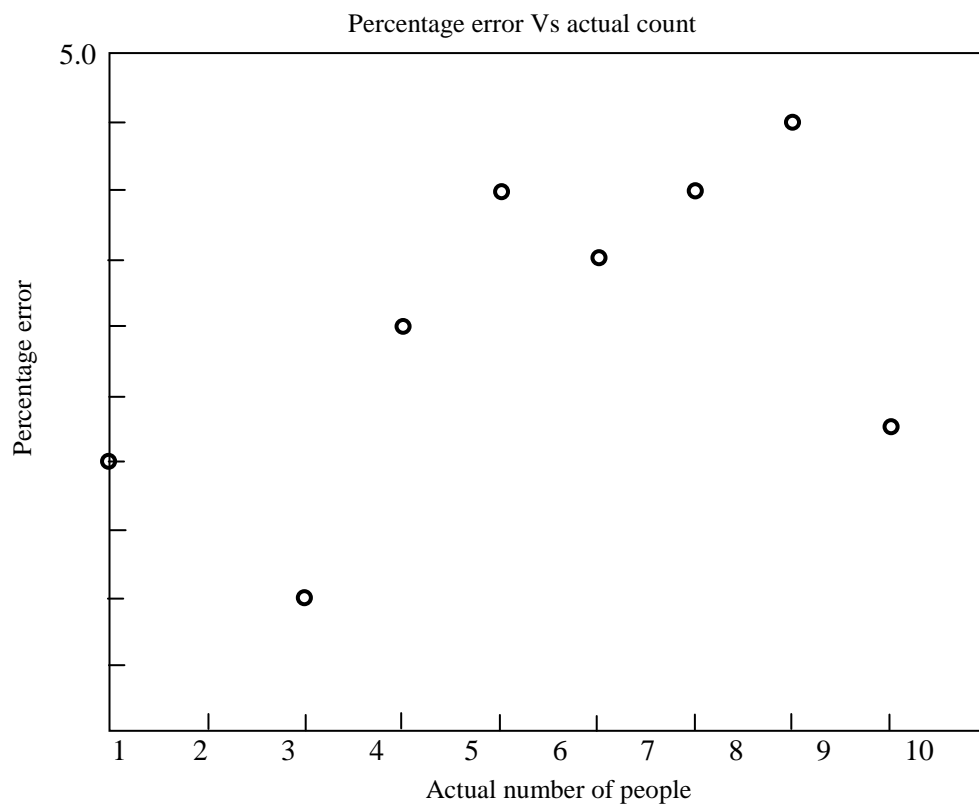


Figure 8. Visual RAM based simulation result comparison.

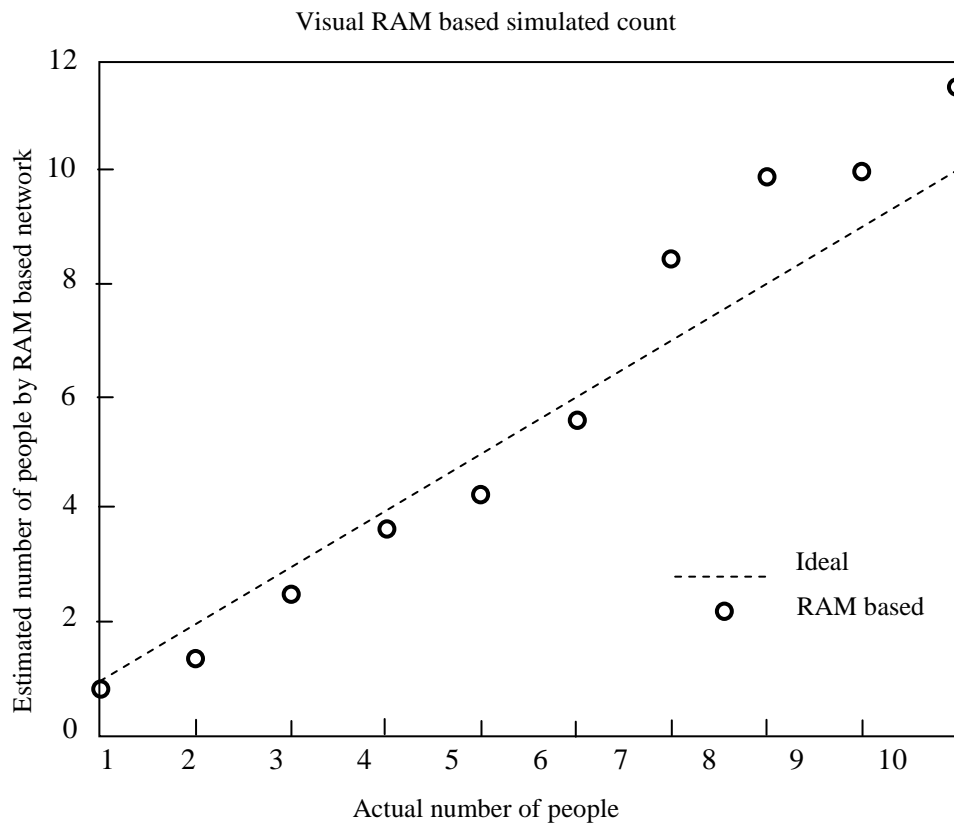


Figure 9. Combined percentage error of visual and infrared systems.

