論文

Studies on Human Skin Extraction from Hyperspectral Data using Particle Swarm Optimization

Particle Swarm Optimization を用いたハイパースペクトルデータの人肌抽出に関する研究

Takayuki EDANAGA, Yohei MINEKAWA,
Sildomar TAKAHASHI MONTEIRO and Yukio KOSUGI
枝長　孝幸*・嶺川　陽平*
モンテイロシウドマールタカハシ*・小杉　幸夫*

Abstract : If the automatic detection of the victims is feasible after large-scale disasters, the search efficiency would be significantly improved compared to the traditional visual search method from aircraft. In order to extract the human skin correctly, we focus on the training of an Artificial Neural Network (ANN) by the hyperspectral data. From the standpoint of avoiding the over-fitting problem generated by multi-channel inputs of the hyperspectral data, it is necessary to be trained by the optimal bands. In this paper, we propose the coupled search method for the optimal number and combination of bands in order to extract the target from the hyperspectral data. The coupled search method is composed of two search methods. The first is a new search method of the maximum value of the evaluation function which searches the optimal number of bands from the non-training data. The other is the search method using the Particle Swarm Optimization which searched the optimal combination of bands from the training data. The ANN is trained by the selected combination of bands, and the results are evaluated. Moreover, the trained network obtained using the coupled search method is compared with the ANN trained by all the bands and the Normalized Difference Human Index.

和文要旨：大規模自然災害の発生後，広域な災害フィールドから人の所在を自動的に探知可能であれば，従来の航空機から目視で被害者を捜索していた効率が向上する。そこで，正確にヒトの人肌抽出を行うために，ハイパースペクトルデータを用いたArtificial Neural Networkの学習に着目する。しかし，ハイパースペクトルデータの全波長を入力として使用した場合，学習データに対してオーバーフィッテングが発生する。本論文では，ハイパースペクトルデータから目標抽出に最適なバンド数と組み合せを探索する手法を提案する。この手法は，未学習データから評価関数の最大値となる最適なバンド数の選択を行い，学習データからParticle Swarm Optimizationを用い最適なバンドの組み合せを探索する。さらに，探索より得られた学習ネットワークは，全波長を入力として使用した学習ネットワークとNormalized Difference Human Indexと比較される。

*東京工業大学大学院総合理工学研究科
Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology

「写真測量とリモートセンシング」VOL. 47, NO. 3, 2008

1. Introduction

Large-scale disasters, such as the tsunami by the Sumatra earthquakes and the flood by Hurricane Katrina which hit the southeast coast of the United States of America, occurred in the world. After such disasters, one of the most important activities is to locate and help the victims as quickly as possible. When relief teams search for victims of large-scale disasters, an automatic human detection system would be of great advantage. Compared to the traditional search-by-eye method from aircraft, the search efficiency would be significantly greater using an automatic detection system. In order to automatically detect the location of victims, we focus on the spectral characteristics of the human skin. As researches for the spectral characteristics of the human skin, the spectral features of the human skin at the wavelength range from 400 to 2000nm have been investigated from the viewpoint of medical applications (Bashkatov, et al., 2005). On the other hand, the identification of individuals using spectral information and the human skin extraction using the spectral absorption properties of hemoglobin in blood vessels have also been studied (Pan, et al., 2003, Angelopoulou, 1999). In addition, we have attempted the human skin extraction using the Normalized Difference Human Index (NDHI) (Edanaga, et al., 2007). The NDHI is calculated from the difference of reflectance between 2 bands in the Near Infrared (NIR) and the Short Wave Infrared (SWIR), which can extract the human skin using a threshold operation. However, it is difficult for the NDHI to deal with some parts of the human skin, such as shaded areas and surfaces that are not parallel to the measuring plane of the sensor due to the effect of the Bidirectional Reflectance Distribution Function (BRDF) (Kimes, 1983). Assuming the adverse condition in the data acquisition from aircraft, the method of the human skin extraction unaffected by partially low light intensities and the measuring angle would be required. If the human skin areas could be extracted correctly, the information would be available to grasp the body shape, the size, and the condition of victims. Thus, we attempt to train an Artificial Neural Network (ANN) by the hyperspectral data. The ANN of a Multilayer Perceptron (MLP) model is suitable for combining multi-channel spectral components nonlinearly. In the combinatory calculation, the accuracy of the human skin extraction is expected to be enhanced by providing a greater number of bands to the ANN as inputs. However, due to the fact that the hyperspectral data is composed of much redundant information (Lillesand and Kiefer, 2000), an over-fitting to the training data occurs when too many bands are provided to the ANN as inputs. This problem is caused by increasing the number of the weights from input layer to hidden layer in the ANN, and too much freedom of the weights is inevitably introduced. Therefore, it is necessary to search for the optimal number of bands in order to avoid the over-fitting problem, and the optimal combination of bands must extract the human skin correctly.

In order to improve the general versatility of the ANN by reducing the redundant spectral components, various methods have been studied for selecting the significant bands which are applicable to hyperspectral data (Kumar, et al., 2001, DeBacker, et al., 2005, Minekawa, et al., 2007). As the search method for selecting the optimal combination of bands, we focus on the Particle Swarm Optimization (PSO). The PSO is one of the optimization algorithms proposed by Kennedy and Eberhart in 1995 (Kennedy and Eberhart, 1995). The PSO searches for an optimal solution by exchanging the information of each particle's position. In recent years, the PSO has received much attention as a useful search method for an optimal solution over a continuous space. Since the adjustment of the PSO's parameters is relatively simple, the PSO has been studied for use in various applications (Eberhart and Hu., 1999, Yoshida, et al., 2000). As the previous study of hyperspectral data analysis using the PSO, the search method for the optimal number and combination of bands in order to presume glucose content in soybeans has been studied (Monteiro and Kosugi, 2007). In this search method, the Penalty Function (PF) and the Performance Evaluation Function (PEF) are used for the selection of the searched number of bands. However, due to the fact that the operator judges the optimal number of bands and adjusts parameters of the PF and the PEF, there is no assurance that the searched number of bands is the optimum. In this paper, we propose a new method in order to search for the optimal number of bands which is able to avoid the over-fitting problem and extract the targets from the hyperspectral data correctly. We then try to couple this method with another PSO framework, and execute the coupled search method for the basic analysis of the human skin extraction.

This paper is organized as follows. In section 2, a new method to search for the optimal number of bands is explained. In addition, the components of the PSO framework are explained. In section 3, experimental results with the optimal number and combination of bands for the human skin extraction are shown and validated. In section 4, the extraction method using this coupled search method is compared with that using the ANN which is trained by all the bands and the NDHI, and the reliabilities of this coupled search method are evaluated.

2. Method for Analysis

First, the evaluation function used in the search for the number of bands is introduced. The sizes of the human skin and the background are different within the acquired hyperspectral data. In particular, assuming a case of the search for victims who are locating over a large disaster field, it is clear that the size of any visible human skin is significantly smaller compared to the size of the background. The averaged accuracy rate (AR_{AV}) is defined in equation (1) to evaluate the human skin and the background as the same weights:

$$AR_{AV}=\frac{1}{2}(AR_H+AR_B), \tag{1}$$

where AR_H is the accuracy rate of the human skin extraction, AR_B is the accuracy rate of the background extraction. The AR_H and the AR_B are calculated in the following:

$$AR_H=\frac{\sum_{n=1}^{N_H}H_k}{N_H},\ AR_B=\frac{\sum_{n=1}^{N_B}(1-B_k)}{N_B}, \tag{2}$$

where N_H is the pixel number of the human skin areas, N_B is the pixel number of the background areas, H_k is the normalized network output for each pixel in the human skin areas ($0\leq H_k\leq 1$), B_k is the normalized network output for each pixel in the background areas ($0\leq B_k\leq 1$).

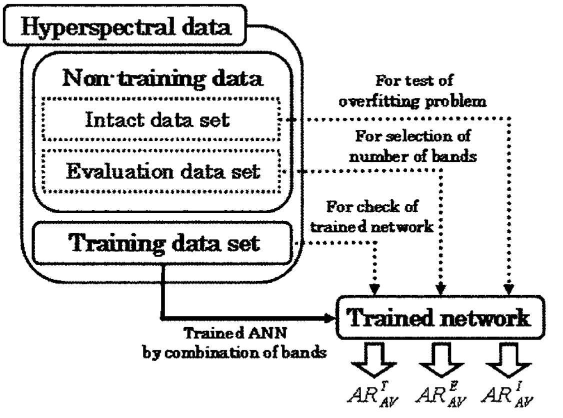


Fig. 1 Type of the hyperspectral data set.

Next, the hyperspectral data used in this paper is divided into three data sets: training data set, evaluation data set and intact data set as shown in Fig. 1. The training data set is utilized to train the ANN and search for the optimal combination of bands using the PSO. In the non-training data section, the evaluation data set is employed to search for the optimal number of bands, and the intact data set is employed to test the over-fitting problem. The number and combination of bands are searched from the training data set and the evaluation data set. The trained network learns by the training data set. The AR_{AV} is calculated by applying each data set to the trained network: the AR^T_{AV} of the training data set, the AR^E_{AV} of the evaluation data set, the AR^I_{AV} of the intact data set. Whether the over-fitting problem occurs or not is tested by applying the intact data set to the trained network.

In order to search for the optimal number of bands to extract targets from the hyperspectral data, we focus on the effects of the number of bands on the over-fitting. Fig. 2 shows the variation of the AR^T_{AV} and the AR^E_{AV} in the case where the number of bands is increased. When the number of bands is increased, the AR^T_{AV} is generally increased until the number of bands reaches the maximum number. In contrast, when it exceeds the specific number of bands, the AR^E_{AV} is decreased as the increment of the number of bands due to the over-fitting problem. Therefore, we propose the search method of the maximum AR_{AV}^{E} which searches the optimal number of bands. This method searches the maximum AR_{AV}^{E} and judges its number of bands as the optimal one. By using this method, the trained network is expected to have the general versatility against the evaluation data set and avoid the over-fitting with respect to the training data set. By increasing the types of the evaluation data set significantly, it becomes possible to find the number and combination of bands which possess high general versatility against the non-training data.

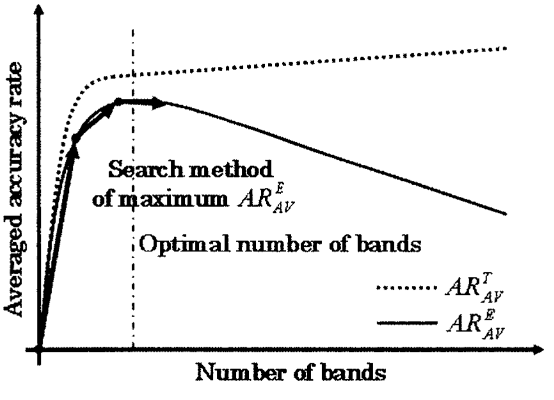


Fig. 2 Schematic illustration of the variation of the AR_{AV}^{T} and the AR_{AV}^{E} in cases where the number of bands is increased, and the optimal number of bands is searched using the search method of the maximum AR_{AV}^{E}.

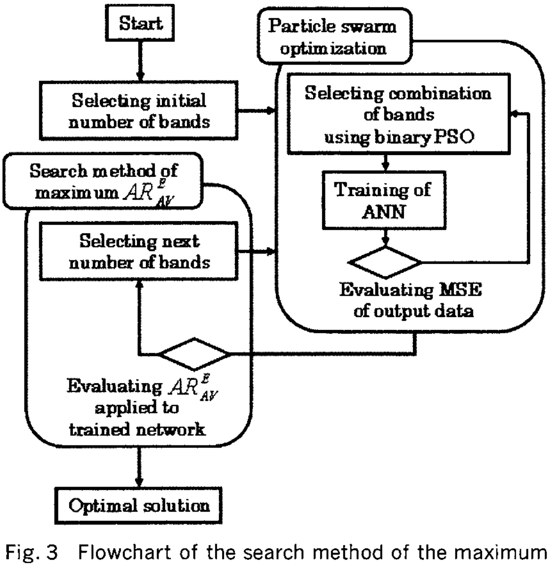


Fig. 3 Flowchart of the search method of the maximum AR_{AV}^{E} coupled with the PSO.

The outline of the search method of the maximum AR_{AV}^{E} coupled with the PSO is explained as follows. Fig. 3 is a flowchart of the coupled search method. Firstly, the initial number of bands is selected. Next, the PSO searches various combinations of bands. Binary PSO (Agrafiotis and Cedeño, 2002) is used to select the bands. Then the ANN is trained by the selected combination of bands. Evaluating the Mean Squared Error (MSE) of the trained network output, the PSO repeats the search for the optimal combination of bands. If the MSE does not change in the pre-determined iterations or the number of the PSO's iterations reaches the pre-determined maximum number, the trained network is output as that which is trained by the optimal combination of bands for the selected number of bands. Next, in order to evaluate the selected number of bands, the AR_{AV}^{E} is calculated using this trained network. The search method of the maximum AR_{AV}^{E} determines whether the calculated AR_{AV}^{E} is the maximum or not. If the AR_{AV}^{E} is the maximum, the selected number and combination of bands are the optimal ones. If not, the subsequently searched number of bands for the PSO's search is determined, and the process is repeated. In the following, the search method of the maximum AR_{AV}^{E}, PSO, binary PSO, and ANN are described.

2.1 Search method of the maximum AR_{AV}^{E}

In the search method of the maximum AR_{AV}^{E}, the optimal number of bands is searched from the AR_{AV} of the evaluation data set. This method searches for the number of bands in the order of the number from smallest to largest by a pre-determined step number of bands. At first, the AR_{AV}^{E} is increased as the increment of the searched number of bands. Then, the trajectory of the search points in the range where the maximum AR_{AV}^{E} exists is convex which is similar to that of the search points in Fig. 4 (1 → 2 → 3). The prospective trajectories can be classified into three categories: the chain line, the bold line, and the thin line as shown in Fig. 4. After the convex trajectory is found, the following search procedure

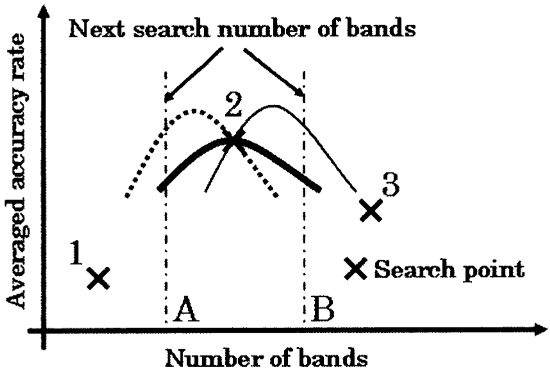


Fig. 4 Schematic illustration of the algorithm which can search the maximum AR^{E}_{AV} in the range where the trajectory of the search points is convex.

is executed in order to search for the maximum AR^{E}_{AV} in the range.

[Search procedure]

i) After the convex trajectory of the search points is found, this range is defined as the search range. The search points are made up from the small to the large number of bands in the search range, and each number of bands is defined as the search point 1, 2, and 3.

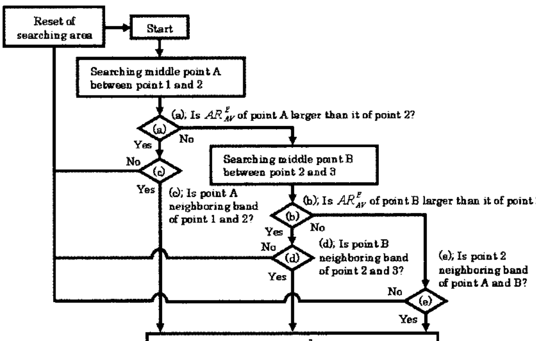
ii) The middle point "A" between the search points 1 and 2 is searched. If the AR^{E}_{AV} of the middle point "A" is smaller than that of the search point 2, the process goes to iii). If the AR^{E}_{AV} of the middle point "A" is larger than that of the search point 2, the maximum point exists in the range between the search points 1 and 2 (the trajectory of 1 → A → 2 is convex). Moreover, if the number of bands of the middle point "A" is located next to that of both the search points 1 and 2, the maximum point is determined to be the middle point "A". If not, the process goes back to i).

iii) The middle point "B" between the search points 2 and 3 is searched. If the AR^{E}_{AV} of the middle point "B" is smaller than that of the search point 2, the process goes to iv). If the AR^{E}_{AV} of the middle point "B" is larger than that of the search point 2, the maximum point exists in the range between the search points 2 and 3 (the trajectory of 2 → B → 3 is convex). Moreover, if the number of bands of the middle point "B" is located next to that of both the search points 2 and 3, the maximum point is determined to be the middle point "B". If not, the process goes back to i).

iv) When the AR^{E}_{AV} of the middle points "A" and "B" are smaller than that of the search point 2, the maximum point exists in the range between the middle points "A" and "B" (the trajectory of A → 2 → B is convex). If the search point 2 is located next to that of both the middle points "A" and "B", the maximum point is determined to be the search point 2. If not, the process goes back to i).

*The number of significant figures for this calculation is 14 or 15. Due to the very small number of significant figures, each AR^{E}_{AV} is not the same value in most cases. In order to simplify the search procedure, the search is judged as a failure in the case where each AR^{E}_{AV} is the same value.

Fig. 5 is the flowchart of the algorithm which can search the maximum AR^{E}_{AV} in the range where the trajectory of the search points is convex. If we use the variation of the gradient in order to search for the maximum AR^{E}_{AV}, it is difficult to adjust the search parameters when the variation of the gradient becomes very small partially. In the proposed method, only one parameter for the step number of bands is necessary. Therefore the determination of the parameter is easy. In this work, we try to execute the search algorithm with the condition that the selected step number of bands is 5.



*If AR^{E}_{AV} of middle point A or B is equal to it of point 2 in (a) and (b), searching is failed.

Fig. 5 Flowchart of the algorithm which can search the maximum AR^{E}_{AV} in the range where the trajectory of the search points is convex.

2.2 Particle Swarm Optimization

After selecting the number of bands, the optimal combination of bands is searched using the PSO algorithm. The PSO algorithm is briefly described in the following. At first, all particles are arranged in random positions on the feature space which is parameterized for the problem. Next, each particle exchanges information of their position vector with other particles, and update the velocity vector and the position vector of one's own. Thus, each particle moves on the feature space. The optimal combination of bands is searched by repeating this operation. The velocity vector of the PSO algorithm (Shi and Eberhart, 1998) is calculated using the following equation :

$$v_{id}^{t+1} = \omega v_{id}^{t} + c_1 r_1 (p_{id}^{t} - x_{id}^{t}) + c_2 r_2 (p_{gd}^{t} - x_{id}^{t}), \tag{3}$$

where c_1 and c_2 are referred to weights which determine the relative influence of the best position of the individual record versus that of the best position among all the particles, r_1 and r_2 are random numbers in the range [0, 1], ω is an inertia weight, $X_i = (x_{i1}, x_{i2}, \ldots, x_{iD})$ represents the position vector of the i^{th} particle which is treated as a point in D dimensional inputs, $P_i = (p_{i1}, p_{i2}, \ldots, p_{iD})$ represents the best position vector of the individual record, $V_i = (v_{i1}, v_{i2}, \ldots, v_{iD})$ represents the velocity vector for the i^{th} particle, t indicates the iteration number, g indicates the index of the best particle among all the particles in the population. If the right side of equation (3) exceeds the specified value, particle velocities in each dimension are truncated to the maximum velocity. The position vector is calculated using the following equation :

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1}, \tag{4}$$

The update of the position vectors and the velocity vectors are schematically explained in Fig. 6. Utilizing equations (3) and (4), particles update the velocity vector and the position vector of one's own. By repeating the movement on the feature space, particles gather to the optimal position. The parameters for the PSO are shown in Table 1.

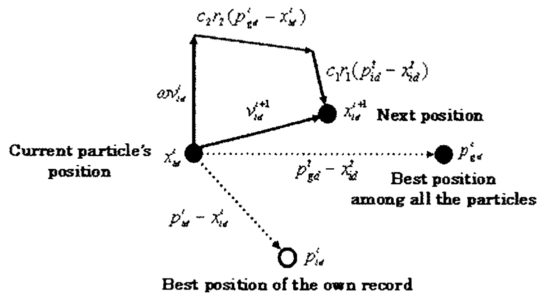


Fig. 6 Update of the particle's position vector and velocity vector.

Table 1 Parameters for the PSO.

Parameter	Value
Population size binary particle swarm	40
Learning rate $C_1 = C_2$	2
Maximum particle velocity	4
Maximum number of epochs	200
Maximum epochs with constant error	30
Initial inertia	0.9
Final inertia	0.2
Epoch of final inertia	190

2.3 Binary PSO

In order to select the band using the PSO, the framework of the binary particle swarms is introduced (Agrafiotis, and Cedeño, 2002). This framework transforms each particle's position vector into a probability of a roulette wheel. The probability is used as the candidate of the band selection. The band is selected by spinning this roulette wheel. The selected band is eliminated from the candidates of the band selection, and the roulette wheel is reassembled and executed again. This process is repeated until the necessary number of bands has been selected. The probability of the roulette wheel is calculated in the following equation :

$$P_r(x_{id}) = \frac{x_{id}^{\ \alpha}}{\sum_{d=1}^{n} x_{id}^{\ \alpha}}, \tag{5}$$

where x_{id} is the particle's position vector, n is the number of all the bands which are acquired from the hyperspectral sensor ($n = 149$ bands), α is a scale factor which controls the probability of the band selection ($a=1$), P_r indicates the probability of the band selection that is represented by the dimension d of the i^{th} particle.

2.4 Artificial Neural Network

The ANN is trained by a selected combination of bands. The MLP network of the nonlinear model is composed of an input layer, a hidden layer, and an output layer connected in a feed-forward way (Haykin, 1999). The activation function for the hidden layer is a hyperbolic tangent sigmoid transfer function. The output function for the output layer is a linear transfer function. The employed training method is the Levenberg-Marquardt back-propagation method (Hagan and Menhaj, 1994). The training data set is composed of the training parts and the validation parts. The training parts (200 points of the human skin areas and 200 points of the background areas) are selected randomly from the training data set. The validation parts (100 points of the human skin areas and 100 points of the background areas) are selected randomly from the training data set excepting the training parts. When the ANN is trained, the MSE of the validation parts is checked in order to avoid the over-fitting. The training of the ANN is terminated in the case where the MES of the validation parts is continuously increased over five times. The ANN is trained in order to decrease the MSE of the training parts. The MSE is calculated using the following equation :

$$MSE(y)=\frac{1}{N}\sum_{i=1}^{N}(y^0-y^t)^2, \tag{6}$$

where y^0 is the target data, y^t is the output data. In this work, the trained network with the smallest MSE within three trials is adopted. The parameters for training of the ANN are shown in Table 2.

Table 2 Parameters for training of the ANN.

Parameter	Value
Input layer size*	1-149
Hidden layer size	10
Output layer size	1
Maximum number of epochs	2000

*Input layer size is changed by selected band number.

3. Experiment

3.1 Experimental Equipment and Data Acquisition

The hyperspectral line sensor is composed of a spectral apparatus (Imspector-N17 from Specim) and a CCD array (SU320M-1.7RT by Sensors Unlimited). This sensor can acquire the spectral data between 950nm and 1700nm wavelengths along the sensing line on target objects. The spectral resolution is approximately 5nm. The spatial resolution along the sensing line is 240 pixels. This sensor is fixed on a mega-torque motor (YSB2020 by NSK) in order to scan the 2-dimensional field. By rotating the motor, the sensor is able to scan in a horizontal direction, and a 2-dimensional hyperspectral image is obtained.

The hyperspectral data including the human skin and other surfaces of the background are acquired at the outdoor. The ethnic group of the examinees is classed as Japanese. In order to confirm the performance of the coupled search method by way of the basic analysis of the human skin extraction, the background objects are simple compositions, i.e. a concrete wall, an asphalt road, and vegetation as shown in Fig. 7. When the hyperspectral data are acquired, the raw data are calibrated with respect to the digital number observed from a standard white board after deducting the dark current value in order to yield the reflectance. The reflectance is calculated using the following equation :

Fig. 7 Background objects (a concrete wall, an asphalt road and vegetation).

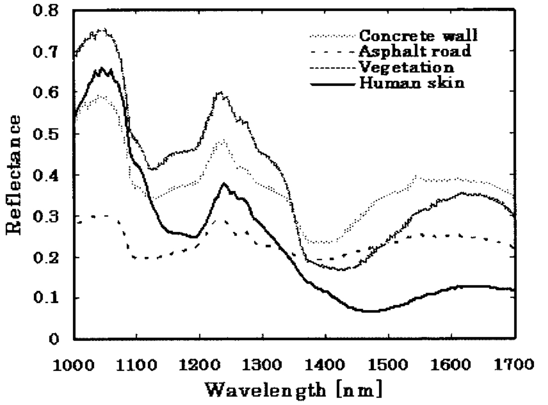


Fig. 8 Hyperspectral reflectance profiles of a human skin, a concrete wall, an asphalt road and vegetation.

$$R(\lambda_j)=\frac{Is(\lambda_j)-Id(\lambda_j)}{Iw(\lambda_j)-Id(\lambda_j)}, \tag{7}$$

where $R(\lambda_j)$ is the reflectance, and λ_j is the j^{th} wavelength ($j=1...n$), $Is(\lambda_j)$ is the digital number of objects which is measured by the hyperspectral sensor, $Iw(\lambda_j)$ is the digital number of the standard white board which is measured by the hyperspectral sensor, $Id(\lambda_j)$ is the dark current value. Fig. 8 shows the hyperspectral reflectance profiles of the human skin and each of the background objects. Three different individuals are included in each data set.

3.2 Analysis Results

The optimal number and combination of bands are searched using the search method of the maximum AR^E_{AV} coupled with the PSO. When the combination of bands is searched, the PSO's search is repeated five times for the purpose of improving the accuracy and validating the AR^E_{AV}. In the five trials, the trained network which outputs the maximum AR^E_{AV} is adopted as that of the selected number of bands.

The AR^E_{AV} and the transfer pathway ($1 \rightarrow 7$) of the search points are shown in Fig. 9. The AR^T_{AV} is shown in Fig. 10. The AR^T_{AV} is sufficiently high when the number of bands reaches 6. Therefore, the AR^T_{AV} is not improved significantly if the number of bands is increased. In contrast, the AR^E_{AV} is increased up to 9 bands, and the maximum AR_{AV} reaches 0.9784.

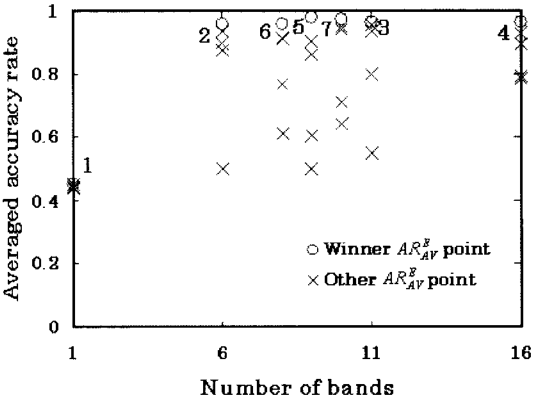


Fig. 9 $\boldsymbol{AR^E_{AV}}$ and the transfer pathway ($1 \rightarrow 7$) of the search points in the search process.

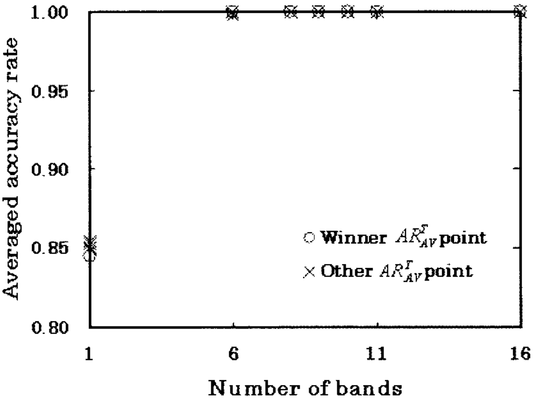
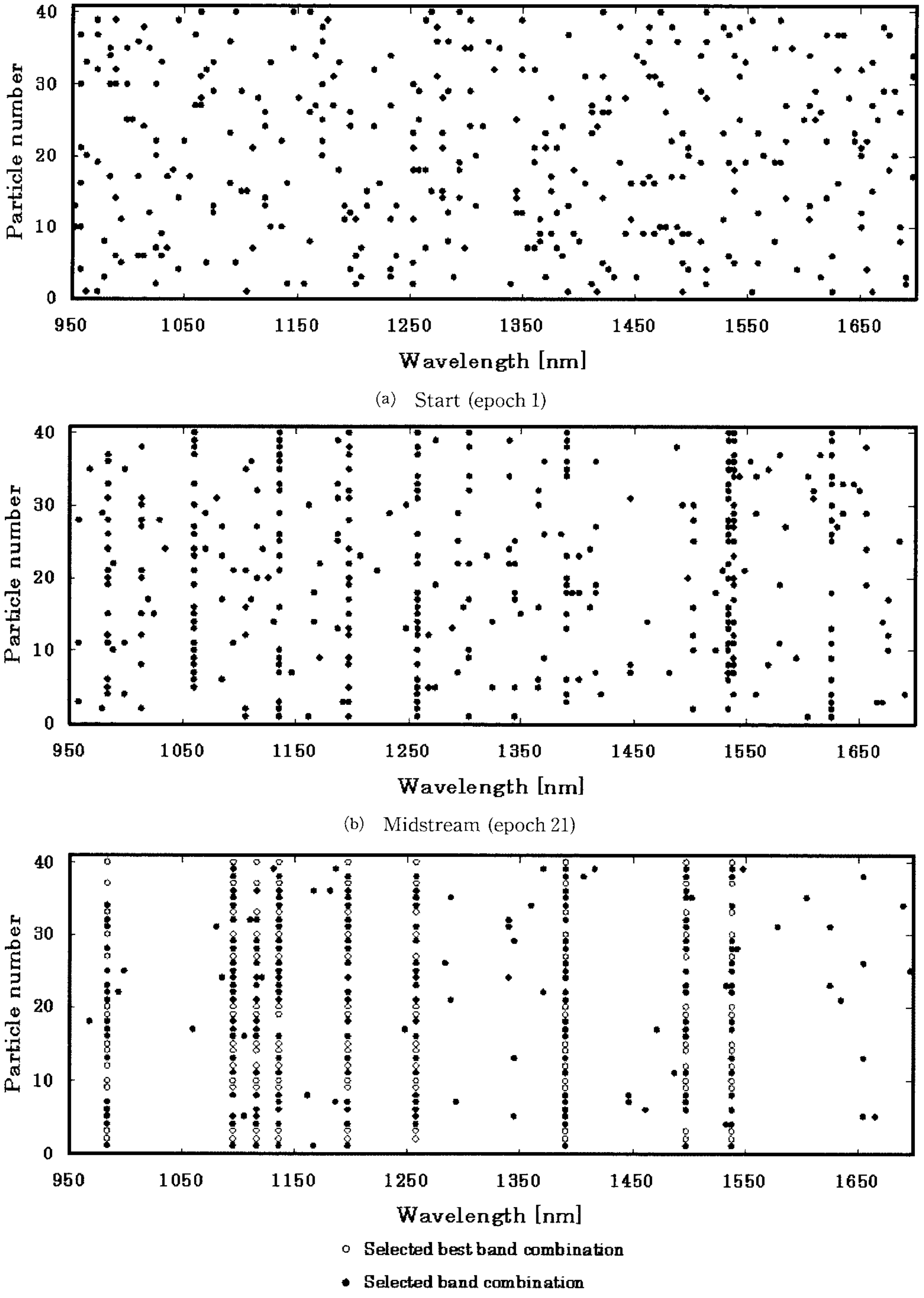


Fig. 10 $\boldsymbol{AR^T_{AV}}$ of the search points in the search process.

After 9 bands, the AR^E_{AV} is gradually decreased. As a result, it is possible to determine the optimal number of bands meeting condition that the AR^T_{AV} is nearly equal to 1 and the AR^E_{AV} is a maximum. The total number of searched bands using the search method of the maximum AR^E_{AV} is 7.

Each AR^E_{AV} obtained from five PSO's trials in the selected number of bands is different. Thus, the search process of PSO's particles is examined in more detail. Fig. 11 show the movements of PSO's particles at 9 bands which gives the maximum AR^E_{AV}. At the beginning step (Fig. 11(a)), 40 particles select the bands randomly. At the intermediate step (Fig. 11(b)), the particles are searching the optimal combination of bands and beginning to select the specified bands. At the final step (Fig. 11(c)), a total of 14 particles select the optimal combination of



(a) Start (epoch 1)

(b) Midstream (epoch 21)

(c) Final (epoch 47)

Fig. 11 Movements of PSO's particles at 9 bands which gives the maximum AR_{AV}^{E}.

bands. As a trend of the all particle's selection, most particles select the bands which are included in the optimal combination of bands. Consequently, the PSO search performs well in itself. Next, the reproducibility of the ANN is verified. By repeating the training of the ANN under the condition which

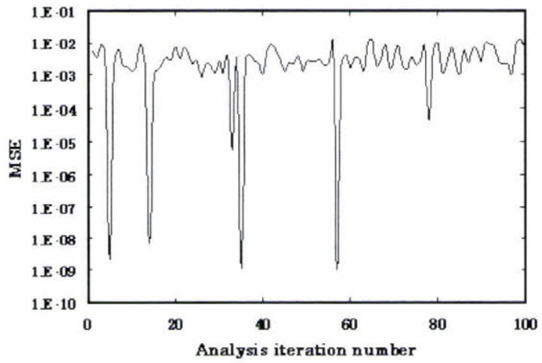


Fig. 12 Variations of the MSE by repeating the training of the ANN under the condition which gives the maximum AR^E_{AV}.

Fig. 13 Grayscale image at 1100nm wavelength (the evaluation data set).

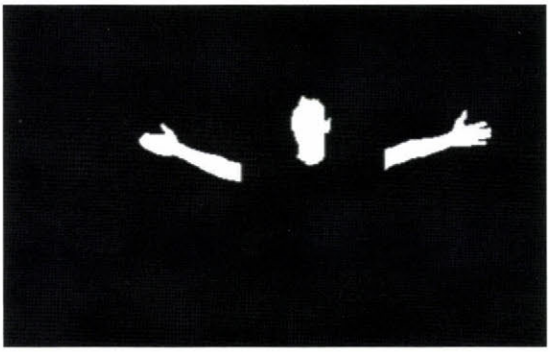


Fig. 14 Grayscale image of the human skin extraction using the trained network with the optimal bands.

gives the maximum AR^E_{AV}, the variations of the MSE are shown in Fig. 12. Most of the MSE in one hundred training trails are in the order of 10^{-3}. However, there are a few cases where the MSE are in the order of 10^{-9}. Furthermore, the minimum MSE in the PSO's search is 1.647×10^{-13}. Therefore, the MSE output from the ANN is not consistent. The fluctuation of the MSE affects on the AR^E_{AV} significantly. As a result, the AR^E_{AV} are poorly reproducible. This problem is due to using the ANN of the MLP model. When a single-layer perceptron (SLP) model is used, the MSE becomes constant and the AR^E_{AV} is reproducible. However, the MSE from the SLP model is larger compared to the results from the MLP model, and the AR^E_{AV} is low. In the case where the ANN of the SLP model is trained by the optimal bands, the AR^E_{AV} is 0.8616. Therefore, the MLP model is adequate as an ANN model in this work.

At this moment, we define the name of the trained network which outputs the maximum AR^E_{AV} as the trained network with the optimal bands. Next, outputting the images of the human skin extraction, the performance of the trained network with the optimal bands is evaluated. As a reference, the grayscale image of the evaluation data set at 1100nm wavelength is shown in Fig. 13. Fig. 14 is the grayscale image after applying the evaluation data set to the trained network with the optimal bands. In Fig. 14, the human skin areas are clearly extracted from the background. In spite of having the high extraction performance of the human skin, the trained network with the optimal bands does not over-fit with respect to the training data set. Therefore, the coupled search method is effective in the search for the optimal number and combination of bands from the hyperspectral data.

4 . Performance Comparison of Human Skin Extraction

4.1 ANN trained by all bands

Using all the bands of the hyperspectral data as inputs, the ANN is trained to extract the human skin. The conditions and the training data set of the ANN are the same as those used for the training of the ANN with the PSO's search. The ANN is trained by the training data set, and we defined the

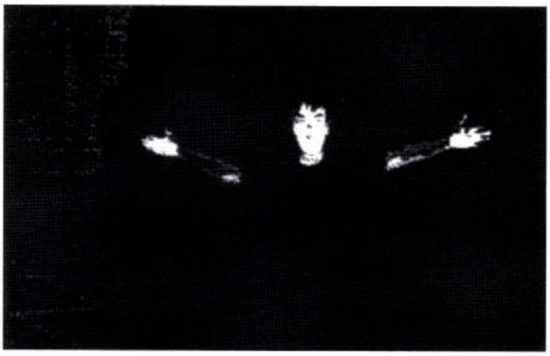


Fig. 15 Grayscale image of the human skin extraction using the trained network with all the bands.

Fig. 16 Grayscale image of the human skin extraction using the NDHI.

name of the output trained network as the trained network with all the bands. The MSE of the trained network with all the bands is 1.018×10^{-13}. Since all the bands are used as inputs, the MSE is smaller than that of the trained network with the optimal bands. It indicates that the trained network with all the bands is well trained by the training data set. Fig. 15 is the grayscale image after applying the evaluation data set to the trained network with all the bands. In Fig. 15, most of the human skin areas are not successfully extracted. The trained network with all the bands generates the over-fitting problem with the evaluation data set. Consequently, the accuracy of the human skin extraction is decreased when the trained network with all the bands is applied to other non-training data.

4.2 Normalized Difference Human Index

The NDHI is an index of the human skin extraction using the spectral characteristics (Edanaga, et al., 2007). In the NIR and the SWIR, the human skin strongly reflects light around 1100nm and absorbs light around 1430nm (Bashkatov, et al., 2005). Thus, the index utilizing the difference of the reflectance at two bands has been proposed. The NDHI has high performances to eliminate the concrete wall and the asphalt road, while it has a comparatively low performance to eliminate the vegetation. The combination of bands which has the highest elimination performance of the vegetation is selected. The NDHI utilizes the difference of the reflectance at 1100nm and 1550nm wavelength. The NDHI is calculated in the following equation :

$$NDHI = \frac{R(1100) - R(1550)}{R(1100) + R(1550)}, \tag{8}$$

where $R(\lambda)$ is the reflectance at λnm wavelength. Fig. 16 is the grayscale image of the evaluation data set output using the NDHI ($0 \leq NDHI \leq 1$). In Fig. 16, the NDHI of the human skin is high. The NDHI of the concrete wall and the asphalt road are low, but that of the vegetation is slightly high.

4.3 Results of performance comparison

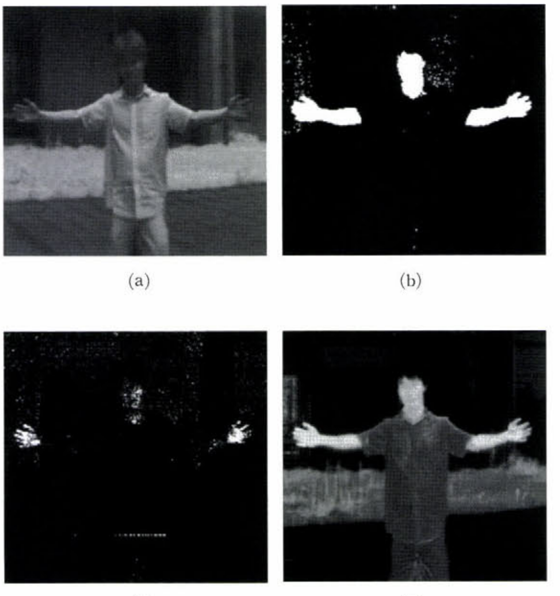
In the extraction performance of the human skin, the trained network with the optimal bands obtained using the coupled search method is compared with the trained network with all the bands and the NDHI. The evaluation data set is used for comparing the extraction performance, and the intact data set is used for testing the over-fitting problem. Table 3 shows the AR^E_{AV} and the percentages of correct pixels after the threshold operation (the threshold value is 0.5) using each of the extraction methods. In the comparison of the AR^E_{AV}, the trained network with the optimal bands has the highest extraction performance. Due to the over-fitting problem, the extraction performance using the trained network with all the bands is low when the other non-training data are applied. The AR^E_{AV} of the NDHI is also low, because the NDHI cannot strongly eliminate the background objects, especially the vegetation. Next, each grayscale image is

Table 3 AR^{E}_{AV} and the percentages of correct pixels after the threshold operation using each of the extraction methods.

	Traind network with all the bands	NDHI	Traind network with the optimal band
AR^{E}_{AV}	0.7241	0.7378	0.9784
Percentage of correct pixels (part of human skin) [%]	44.85	71.24	96.16
Percentage of correct pixels (part of background parts) [%]	99.95	99.95	99.51

transformed into binary images in order to eliminate the background objects, and percentages of correct pixels associated with the human skin areas and the background areas are compared. For the human skin areas, the trained network with the optimal bands is much better than the other extraction methods. The NDHI cannot reliably extract the human skin areas which are shaded or not parallel to the measuring plane of the sensor, and these areas are not extracted after the threshold operation. The trained network with the optimal bands is able to extract these areas sharply. For the background areas, the percentages of correct pixels using each of the extraction methods are almost the same. All of the extraction methods are able to eliminate the background objects after the threshold operation.

Whether the over-fitting problem occurs or not is tested by applying the intact data set to the trained network with the optimal bands. Fig. 17 show the grayscale images of the intact data set using each of the extraction methods, and Table 4 shows the AR^{I}_{AV} and the percentages of correct pixels after the



(a) (b) (c) (d)

Fig. 17 Grayscale image of the intact data set. (a) Grayscale image at 1100nm wavelength. (b) Human skin extraction using the trained network with the optimal bands. (c) Human skin extraction using the trained network with all the bands. (d) Human skin extraction using the NDHI.

Table 4 AR^{I}_{AV} and the percentages of correct pixels after the threshold operation using each of the extraction methods.

	Traind network with all the bands	NDHI	Traind network with the optimal band
AR^{I}_{AV}	0.5874	0.7216	0.9812
Percentage of correct pixels (part of human skin) [%]	16.59	68.30	96.90
Percentage of correct pixels (part of background parts) [%]	99.82	100.00	99.32

threshold operation (the threshold value is 0.5) of Fig. 17(b)(c)(d). The extraction performances of the human skin using the trained network with the optimal bands and the NDHI are as well as the results obtained from the evaluation data set. In the results obtained from two data sets, the trained network with the optimal bands exhibits the highest extraction performance of the human skin. Therefore, the trained network with the optimal bands is effective in avoiding the over-fitting problem that is generated by training of the hyperspectral data.

5. Conclusion

In this paper, we proposed the coupled search method for the optimal number and combination of bands in order to extract the target from the hyperspectral data. The coupled search method was composed of two search methods. The first was a new search method of the maximum AR^{E}_{AV} which searched the optimal number of bands from the evaluation data set which was the non-training data. In this search method, it was possible to search the optimal number of bands which could not have been obtained in the previous study (Monteiro and Kosugi, 2002). In addition, the over-fitting with respect to the training data set was avoided using this method. The other was the search method using the PSO which searched the optimal combination of bands from the training data set. The ANN was trained by the selected combination of bands, and the results were evaluated using the MSE. In general, the greater the size of the training data set is used, the greater the general versatility of the trained network is increased. However, it was difficult to increase the size of the training data set from the standpoint of calculation costs. The coupled search method could be used to consider the evaluation data set in the search for the number of bands. Compared with the training of the ANN by the training data set, the calculation costs for applying the evaluation data set to the trained network were quite low. Hence, it was possible to incorporate a large evaluation data set. According to the results, the coupled search method was able to search for the effective number and combination of bands for the target extraction from hyperspectral data. In the grayscale image of the evaluation data set using the trained network with the optimal bands, the human skin and the background objects were readily separable. Moreover, the extraction performance of the human skin using the trained network with the optimal bands was compared with that using the trained network with all the bands and the NDHI when the evaluation data set was applied. The extraction performance of the human skin using the trained network with all the bands was extremely low because of the over-fitting problem. In contrast, the trained network with the optimal bands could specifically extract the human skin areas where the NDHI could not extract. Therefore, it is expected that the human skin extraction using the trained network with the optimal bands is useful for rescuers to grasp the body shape, the size, and the condition of victims correctly. On the other hand, the trained network with the optimal bands was applied to the intact data set which was not used in the coupled search method, and good results were obtained since the over-fitting problem did not occur.

The following are future works. At the beginning, in order to enable the human skin extraction which is applicable in the various disaster fields, it is required to select the training data set more carefully and increase the evaluation data set effectively. Then, the coupled search method should be evaluated by applying to other extraction targets. In order to simplify problems associated with the coupled search method, the parameters of the PSO and the ANN used in this work were obtained from the previous study. However, there is the possibility that the PSO's particles cannot select the optimal combination of bands on other feature spaces. Hence, it is necessary to optimize the PSO's parameters for each extraction target. In addition, by utilizing other classifiers (e.g. Support Vector

Machine (SVM) (Vapnik, 1998)), the reproducibility and the accuracy of the coupled search method should be improved.

(受付日2008.3.11, 受理日2008.5.9)

References

A.N. Bashkatov, E.A. Genina, V.I. Kochubey, and V. V. Tuchin, 2005. Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000nm. Journal of Physics D: Applied Physics, Vol. 38, pp.2543-2555.

D.K. Agrafiotis and W. Cedeño, 2002. Feature selection for structure-activity correlation using binary particle swarms. Journal of Medicinal Chemistry, vol. 45, pp.1098-1107.

D.S. Kimes, 1983. Dynamics of Directional Reflectance Factor Distributions for Vegetation Canopies. Appl. Opt., vol. 22, no. 9, pp.1364-1372.

E. Angelopoulou, 1999. The Reflectance Spectrum of Human Skin. Technical Report MS-CIS-99-29, pp.1-14, http://www.cis.upenn.edu/~elli/tech-report.skin.pdf.

H. Yoshida, K. Kawata, Y. Fukuyama, S. Takayama, and Y. Nakanishi, 2000. A particle swarm optimization for reactive power and voltage control considering voltage security assessment. IEEE Transactions on Power Systems, vol. 15, no. 4, pp.1232-1239.

J. Kennedy and R.C. Eberhart, 1995. Particle swarm optimization, Proceedings of IEEE international conference on neural networks, vol. 4, pp.1942-1948.

M.T. Hagan and M.B. Menhaj, 1994. Training feedforward networks with the marquardt algorithm, IEEE Transactions on Neural Networks, vol. 5, no. 6, pp.989-993.

R.C. Eberhart and X. Hu., 1999. Human tremor analysis using particle swarm optimization. Proceedings of Congress on Evolutionary Computation, pp.1927-1930.

S. DeBacker, P. Kempeneers, W. Debruyn, and P. Scheunders, 2005. A band selection technique for spectral classification. IEEE geoscience and remote sensing letters, vol. 2, no. 3, pp.319-323.

S. Haykin, Neural Network, 1999. A Comprehensive Foundation, second ed., Prentice-Hall, Englewood Cliffs.

S. Kumar, J. Ghosh, and M.M. Crawford, 2001. Best-bases feature extraction algorithms for classification of hyperspectral data. IEEE transactions on geoscience and remote sensing, vol. 39, no. 7, pp.1368-1379.

S.T. Monteiro and Y. Kosugi, 2007. Particle swarms for feature extraction of hyperspectral data, IEICE trans. inf. & syst., vol. E90-D, no. 7, pp.1038-1046.

T. Edanaga, K. Uto and Y. Kosugi, 2007. Studies on the extraction of human skin using spectral information in shortwave infrared band. Journal of the Japan Society of Photogrammetry and Remote Sensing, vol. 46, no. 2, pp.17-26.

T. Edanaga, K. Uto and Y. Kosugi, 2007. Studies on the extraction of human images in and out of the water using normalized difference human index. 2nd International Conference of GIS/RS in Hydrology, Water Resources and Environment & 2nd International Symposium on Flood Forecasting and Management with GIS and Remote Sensing, vol. 2, 8-pages (CD-ROM).

T.M. Lillesand and R.W. Kiefer, 2000. Remote sensing and image interpretation, John Wiley & Sons, Inc.

V.N. Vapnik, 1998. Statistical learning theory. John Wiley & Sons, Inc.

Y. Minekawa, K. Oda, and Y. Kosugi, 2007. Data Mining Technique in Hyperspectral Remote Sensing. Proc. ICGRHWE2007/FM2S2007, (CDROM 8pages), Guangzhou and Three Gorges.

Y. Shi and R.C. Eberhart, 1998. A modified particle swarm optimizer. Proc. IEEE Congress on Evolutionary Computation, pp.69-73.

Z. Pan, G. Healey, and M. Prasad, 2003. Face recognition in hyperspectral images. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 12, pp.1552-1560.