Korean J. Chem. Eng., **27**(3), 886-892 (2010)
DOI: 10.1007/s11814-010-0139-1

Sludge settleability detection using automated SV30 measurement and comparisons of feature extraction methods

Yejin Kim*, Hoonsik Yeom*, Soojung Choi*, Hyeon Bae, and Changwon Kim*,†**

*Department of Environmental Engineering, Pusan National University, Busan 609-735, Korea
**Busan Techno-Park, Busan, Korea
(*Received 12 August 2009 • accepted 9 October 2009*)

Abstract−The need for automation and measurement technologies to detect the process state has been a driving force in the development of various measurements at wastewater treatment plants. While the number of applications of automation & measurement technologies to the field is increasing, there have only been a few cases where they have been applied to the area of sludge settling. It is not easy to develop an automated operation support system for the detection of sludge settleability due to its site-specific characteristics. To automate the human operator's daily test and diagnosis work on sludge settling, an on-line SV30 measurement was developed and an automated detection algorithm on settleability was developed that imitated heuristics to detect settleability faults. The automated SV30 measurement is based on automatic pumping with a predefined schedule, the image capture of the settling test with a digital camera, and an analysis of the images to detect the settled sludge height. To detect settleability faults such as deflocculation and bulking from these images, two feature extraction methods were used and their performance was evaluated.

Key words: Detection, Diagnosis, Dynamic Time Warping, Discriminant Analysis, Settling, Sludge Volume Index

INTRODUCTION

The application of ICA (instrumentation, control and automation) technologies can be grouped as a process control to ensure optimum performance, process monitoring to get more information from the reactor, and a process diagnosis to detect faults and find their causes. From the field scale implementations of these technologies, it has been proved that well-fitted monitoring and control strategies for the target process can result in an excellent disturbance rejection. However, in most of the ICA system establishment cases, the objectives of applied technologies were concerned with biological removal efficiency management, rather than sludge separation at secondary settling tanks.

It is one of basic concepts of wastewater treatment that sludge settling is exceptionally important to maintain good effluent quality made by activated sludge. With the remarkable progress of measuring techniques including image analysis and the settler models, many variables such as SVI (sludge volume index), SBH (sludge blanket height), effluent turbidity, floc size and the length of filamentous bulking organisms can be measured on-line using automated measurements [18]. However, the operators at field plants check settleability by performing a settling test (SVI test), monitoring SBH at secondary settling tanks, observing the floc size using microscopy, and monitoring effluent solid concentration and turbidity. The reason is that there is much hidden information so that the operators can extract that information from observing an SVI test. While doing the SVI test and obtaining information from that is helpful to the operators to judge settleability qualitatively, it is difficult to automate the test. The SVI test includes various human works such as sludge setting for testing, observing sludge floc size and color, sludge settling velocity and the turbidity of supernatant. Therefore, in most of the cases of automated strategies used to manage effluent SS (suspended solids) quality, the SVI or settling property was excluded or the suggested diagnosis algorithm could not be automated [9,10].

To automatically obtain the SVI value and information concerning the settling state, the automated SVI test was suggested by using automated pumping/draining sludge and obtaining images with a digital camera in this research. This system, the automatic SV30 measurement, consists of a test part which includes pumps and a test cylinder, an image processing part for obtaining and analyzing images from the settling test with a digital camera, and an operating part which controls the on/off function as well as an executing image analysis program. From the images obtained during the automated settling test, the image processing part detects sludge height in the test cylinder and extracts qualitative information of sludge settling such as turbidity of supernatant to detect the sludge settleability problem. To detect the sludge settleability problem, two feature extraction methods were used. The first method involves dynamic time warping, which is a representative pattern matching method, while the second involves discriminant analysis, one of the most generally applied multivariate statistical analysis methods. To develop a fault detection algorithm using these methods, a lab-scale reactor was operated to obtain images of three categories of settling (normal, deflocculation and bulking) under various environmental conditions. The settling fault detection algorithm was then developed and the performance of each method was compared.

MATERIALS AND METHODS

1. On-line SV30 Measurement

1-1. Hardware Structure

The automated SV30 measurement system is equipped with a personal computer in order to save the image file and to perform

†To whom correspondence should be addressed.
E-mail: cwkim@pusan.ac.kr

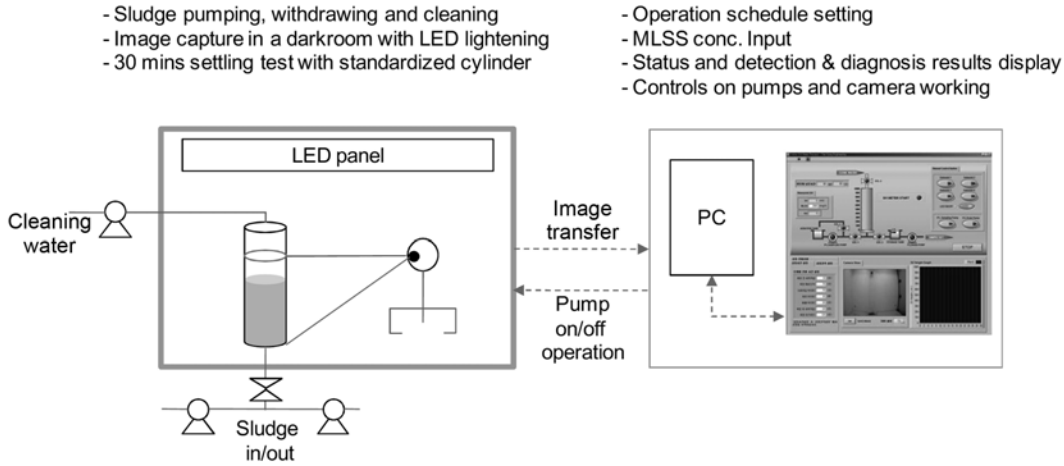

Fig. 1. On-line SV30 measurement system.

image processing, a darkroom containing a test cylinder the size of a standardized 1 L cylinder, a digital camera, and three pumps for sludge pumping/draining and for cleaning water pumping (Fig. 1). In the darkroom, an LED panel was installed as a light source. The on/off operation of all hardware is controlled based on a predefined schedule at an operational program coded with LabVIEW. The steps to perform a fully automated settling test are as follows.

Step 1. Clean cylinder with tap water (2 min)

Step 2. Sludge input to the test cylinder with upflow stream (1 min)

Step 3. Perform settling test and obtain 1 image every 1 minute (31 images, 30 min)

Step 4. Sludge drain (1 min)

Step 5. Clean cylinder with tap water (2 min)

Step 6. Idle to the next measuring

1-2. Image Data Generation & SV30 Detection

Digitalized images consist of pixels and each pixel has its own color information to construct a target subject. Among various color expression methods, RGB is a well known index that explains all colors with a red, green and blue index value from 0 to 255. White has an RGB of (255, 255, 255) and the RGB of black is (0, 0, 0). The colors of settled sludge and supernatant are shown in Fig. 2. After obtaining the image as a pixel size of 390*230 as defined by the resolution of digital camera used, as shown in Fig. 2, one line at the center of the image is extracted and a matrix of 1*230*3 is generated. Only 1*230*1 of the RGB matrix for the R index was extracted and used to detect sludge height in the test cylinder and to identify the settleability problem. The R index was selected to detect the sludge blanket height because the color of sludge is brown, so that the red is most sensitively representing color for the boundary between settled sludge and supernatant. It can be seen in Fig. 2 that the sludge and supernatant part have different color distributions and the sludge height can be detected by determining the knee part of the profile.

To detect the knee part that represents the interface between settled sludge and supernatant, the slope from the value of the bottom (0 ml point, (1, 230) of the R matrix) to each element of the matrix (1, 1-229) was calculated and generated as a new matrix with a size of 1*229. From the newly created slope matrix, the element having

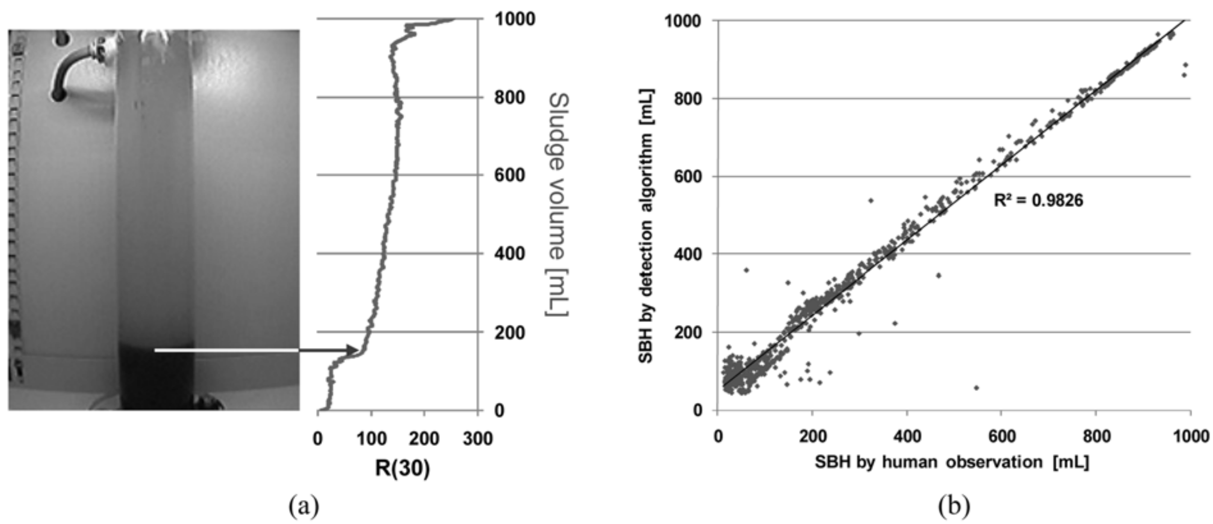

Fig. 2. Description of (a) concept of image analysis and (b) reliability of SBH detection.

maximum value was assumed to be the interface between the solid and the supernatant. The volume information can then be obtained by the conversion of the location of that element to the height. Fig. 2(c) shows the relatively good reliability of this algorithm by comparing the SBH value from human eye observation and the developed algorithm based on 300 points of test results with values of SBH ranging from 50 to 950. After applying to a field-scale WWTP (not included in this paper), the algorithms showed good reliability as it is.

2. Experimental Set-up for Diagnosis of Algorithm Development

To add a settleability problem detecting algorithm to the automated SV30 measurement, various images of the settling problem were collected via a lab-scale aeration tank. The target settling categories are normal settling, deflocculation and bulking [11]. The deflocculation state represents pin-point floc and dispersed growth. The over-aerated condition was created to obtain deflocculation settling (4-6 mg O_2/L). For the bulking case standing for both filamentous and non-filamentous bulking, an extremely low DO (0.1-0.5 mg O_2/L) was given to the reactor. For both experimental cases, the F/M ratio was fixed as 0.18 kg COD/kg MLVSS.d. With the occurrence of each settling problem, a microscopic observation (Zeiss Axiosckop 2plus, Germany, *200) was accompanied to confirm the settling problem with the floc size. The initial normal settling was broken at about three weeks after starting the operation and images were obtained once a day using the automated SV30 measurement.

3. Feature Extraction Methods

The objectives of the research are to develop the automated SV30 measurement system and to provide more clear information about settleability through the automated test with sludge from the aeration tanks. This is for excluding the possible effects of the volume of the settling tank size and operational option such as the sludge underflow rate and the solid loading rate. To obtain information from R profiles, two feature extraction methods were used and their performance was compared.

3-1. Pattern Matching Algorithm

Dynamic time warping (DTW) was introduced by Sakoe and Chiba [12]. DTW has also been used in conjunction with a neural network for the recognition of isolated words. In this study, DTW was used to identify the type of fault among the three cases. The DTW algorithm removes timing differences between R profiles by warping one pattern until it maximally coincides with the other. All pattern vectors are warped against a reference pattern vector of the same category which has the same number of feature vectors as the number of frames in the input layer of the network. After the relevant feature extraction has taken place, each R patterns can be represented as a sequence of feature vectors.

$$A=a_1, a_2, \ldots, a_i, \ldots, a_K$$
$$B=b_1, b_2, \ldots, b_j, \ldots, b_M \quad (1)$$

Let A be the reference R pattern and B be the pattern vector to be aligned against A. Fig. 3 shows A and B developed against the i and j axes.

Consider a warping function F between the input pattern time j and the reference pattern time i, where

$$j=j(i) \quad (2)$$

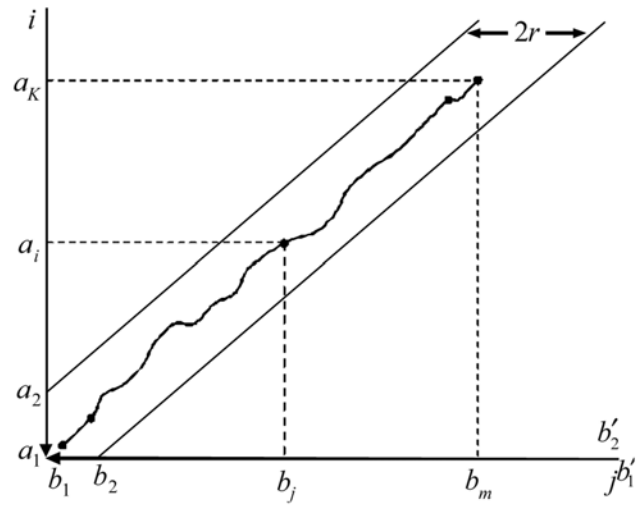

Fig. 3. Warping function and adjustment window.

A measure of the difference between the two feature vectors a_i and b_j is the distance

$$d(i, j)=\|a_i-b_j\| \quad (3)$$

When the warping function is applied to B, this distance becomes

$$d(i, j(i))=\|a_i-b'_j\| \quad (4)$$

where b'_j is the *j*th element of B after the warping function has been applied.

The minimum residual distance between A and B is the distance still remaining between them after minimizing the differences between them. The time normalized difference is defined as follows:

$$D(A,B)=\underset{F}{\mathrm{Min}}\left[\frac{\sum_{i=1}^{K} d(i,j(i))*w(i)}{\sum_{i=1}^{K} w(i)}\right] \quad (5)$$

Applying dynamic programming principles to the simplified time normalization equation gives the following algorithm for calculating the minimal value of the summation: the dynamic programming equations are as follows. All the procedure was coded as MATLAB ver. 7.1.

$$g_i(i, j(i))=\min[g_{i-1}(i-1, j(i-1)+d(i, j(i)*w(i)] \quad (6)$$

$$D(A,B)=\frac{1}{N}g_K(i(K),j(K)) \quad (7)$$

$$g_1(1, 1)=d(1, 1)*w(1)=d(1, 1) \quad (8)$$

$$g_i(i,j)=d(i,j)+\min\begin{cases} g(i-1,j-1) \\ g(i-1,j-2) \\ g(i-1,j-3) \end{cases} \quad (9)$$

3-2. Discriminant Analysis

Discriminant analysis is used for classifying a set of observations into predefined classes. The purpose is to determine the class of an observation based on a set of variables known as predictors or input variables. The model is built based on a set of observations for which the classes are known. This set of observations is sometimes referred to as the training set. Based on the training set, the technique constructs a set of linear functions of the predictors, known as discriminant functions, such as $L=b_1x_1+b_2x_2+\ldots+b_nx_n+c$, where the b′s are discriminant coefficients, the x′s are the input variables or predictors and c is a constant. These discriminant functions are used to predict the class of a new observation with an unknown class. For a k class problem, k discriminant functions are constructed. Given a new observation, all the k discriminant functions are evaluated and the observation is assigned to class i if the i^{th} discriminant function according to the value of functions or rules [13]. To apply the discriminant analysis, SPSS ver. 14.0 was used.

RESULTS

1. Digitalized Information of Settling Problem

The R profiles from images under normal settling, deflocculation, and bulking which are obtained from a lab-scale reactor operation are shown at Fig. 3. The SV30 value at each case was detected from the image and the MLSS concentration was measured by the experimental method to be used for calculating SVI. The R profiles from the images at both the starting and finishing times of the settling test are shown in Fig. 4(a). After 30 min from the start of the test, the settled sludge height was reflected as the knee part of the R profile. Compared to the R profiles of normal settling (Fig. 4(b)), the R profiles of deflocculation (Fig. 4(c)) and bulking (Fig. 4(d)) have distinctly different patterns, which are due to the high turbidity of supernatant and unsettled sludge, respectively.

2. Test and Validation Data

Table 1 shows the data characteristics of test and validation data to be used in the two feature extraction methods. The test data sets were used to generate standard profiles of each settling property at dynamic time warping algorithm and to determine discriminant functions for discriminant analysis. The validation data sets were used to investigate the accuracy of each algorithm. The MLSS concentration was distributed from 1,380 mg/L to 2,930 mg/L during the operation of the lab-scale reactor, as shown at Table 1. The test and validation data sets were obtained in the order of: Normal (Test) - Deflocculation (Test) - Normal (Validation) - Bulking (Test, Validation) - Deflocculation (Validation).

3. Information Extraction with Pattern Matching Algorithm

3-1. Standard Patterns Generation

Classifying a new case to pre-defined groups using pattern match-

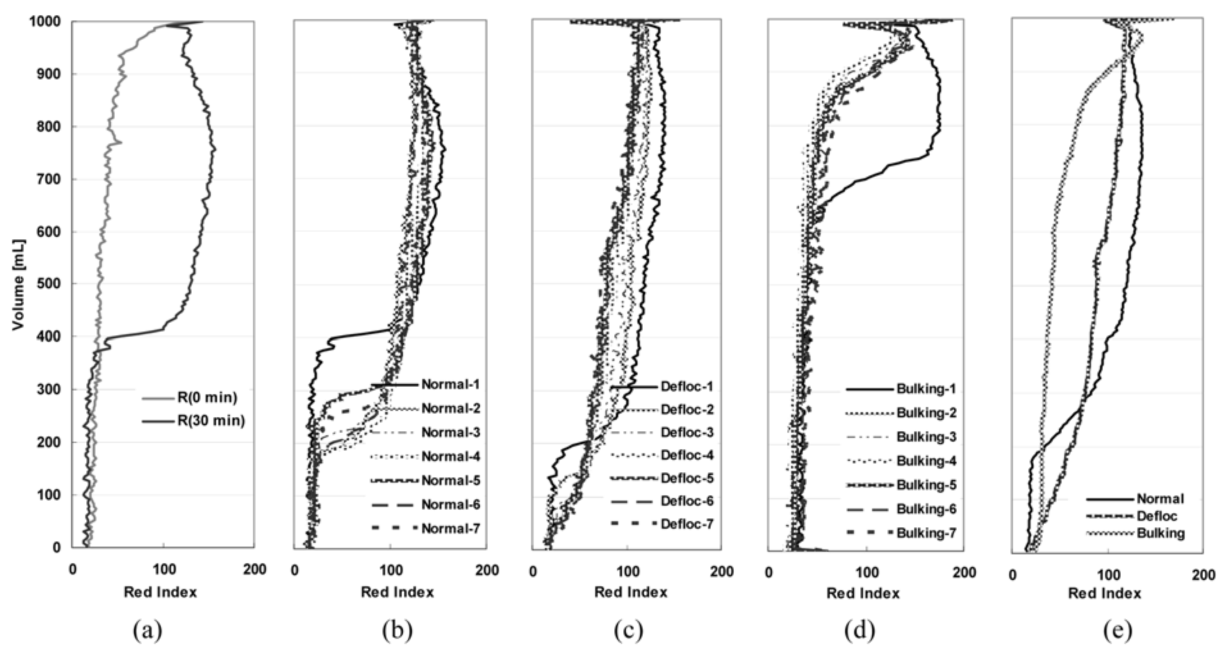

Fig. 4. R profiles from initial and final state of settling test (a), normal settling (b), deflocculation (c) bulking (d) and standard patterns of each fault (e).

Table 1. Characteristics of test and validation data from experimental results

	Normal		Deflocculation		Bulking	
	Test	Validation	Test	Validation	Test	Validation
Set	7	3	10	4	10	5
MLSS	2542.0	2366.67	1818.57	1490.0	2134.0	2190.0
SV30	255.7	358.0	70.4	74.3	936.2	963.8
SVI	102.9	153.4	29.3	50.3	451.0	460.4
Data name	N1-N7	V_N1-V_N3	D1-D10	V_D1-V_D4	B1-B10	V_B1-V_B6

ing is based on comparison between standard patterns of each group and the pattern of a new case. Fig. 4 shows the test patterns of each case which are digitalized from the captured images. For the deflocculation case, the R profiles show shifting from right to left, which means the supernatant part started to have unsettled flocs and have some color. The bending points indicating settled sludge blanket height became indistinct at the case of deflocculation and those are located at the upper part of the cylinder for the bulking because of higher sludge volume and low density. To implement pattern matching, three standard patterns were needed for each settleability category. The standard patterns were generated as average values of test patterns (7 test cases for the normal reference pattern; 10 test cases for the deflocculation; 10 test cases for bulking) (Fig. 4(e)).

3-2. Time Warping with Each Test Data

From the results of pattern matching with a dynamic time warping algorithm, Table 2 shows the D values of test data sets to the three standard patterns. When a test pattern is matched to three standard patterns and has three D values, the D value should be the lowest within that standard pattern, which represents the same group as the test pattern. As shown in Table 2, the D values are smallest at each representative standard pattern except N2 and D1. As shown in Fig. 4(c), D1 can be regarded as an intermittent state between normal settling and deflocculation. Though N2 was a normal state, D_st(D) was smaller than D_st(N) and D_st(B), indicating a weak point in the dynamic time warping application, which is caused by the purity of the generated standard patterns or test data sets. In other words, N2 and D1 need to be identified again to determine which group they should be included in, or the suitability of the standard patterns should be rechecked. In spite of these weak points, the result of the suggested algorithm is extremely simple, whereby a test pattern's group is identified as a group with the smallest D value from dynamic time warping.

4. Diagnosis with Discriminant Analysis

The basic concept of discriminant analysis involves the generation of discriminant functions as a linear combination of independent variables and identifying groups of newly provided case data according to the score of the discriminant functions. In this research, independent variables needed to be extracted in order to present a special feature for three settling properties from the R profiles. Based on the observation of R profiles, different features between an initial R profile (t=0 min) and a final R profile (t=30 min) were targeted to extract variables. An average value of the R profiles for the final R profile (named "R average"), a sum of the difference between the two R profiles (named "R sum") and a standard variation of the final profile (named "R stdev") were generated. These variables include information about the breadth of the profile and the inclina-

Table 2. Results of dynamic time warping for three standard patterns

	N1	N2	N3	N3	N5	N6	N7			
D_st(N)*	**778.7**	701.7	**252.1**	**241.4**	**319.6**	**570.6**	**239.3**			
D_st(D)**	2957.9	**679.1**	1611.1	1231.3	1964.3	741.7	1600.3			
D_st(B)***	4424.1	3874.6	4586.5	4448.5	4329.2	3905.6	4287.7			
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
D_st(N)	**329.1**	845.1	1380.6	2694.6	3172.9	2677.4	3261.3	3482.3	2477.3	2675.3
D_st(D)	1467.1	**477.3**	**245.7**	**202.5**	**309.7**	**189.7**	**301.5**	**326.9**	**202.9**	**201.3**
D_st(B)	4457.7	3578.0	3050.5	2246.1	1884.3	2098.5	1532.2	1669.4	2365.0	2603.9
	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
D_st(N)	2812.6	2744.7	2819.9	3291.3	2680.6	2579.6	2564.7	2736.3	3064.4	4695.7
D_st(D)	3047.3	2689.1	2791.1	3041.1	2490.3	1924.1	1746.1	2152.1	2840.3	2453.1
D_st(B)	**2024.5**	**265.0**	**271.4**	**360.6**	**214.1**	**262.3**	**286.6**	**238.4**	**259.9**	**393.3**

*: The standard pattern of normal settling; **: The standard pattern of deflocculation; ***: The standard pattern of bulking

Table 3. Discriminant analysis results for test data

Normal	N1	N2	N3	N3	N5	N6	N7			
Y1	2.95	5.50	6.53	6.83	4.12	4.75	4.51			
Y2	3.96	0.87	1.49	0.88	1.54	0.39	1.13			
Group	N	N	N	N	N	N	N			
Defloccu-lation	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Y1	**5.03**	4.05	3.96	2.96	2.54	2.96	1.43	1.50	2.58	2.93
Y2	**0.03**	−0.88	−1.14	−1.23	−0.89	−1.12	−1.30	−1.61	−2.10	−2.38
Group	**N**	D	D	D	D	D	D	D	D	D
Bulking	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
Y1	−5.34	−6.03	−5.67	−5.62	−6.23	−6.46	−6.74	−6.44	−5.80	−5.13
Y2	2.20	0.89	0.83	1.06	0.66	−0.70	−0.82	−0.44	0.82	−0.40
Group	B	B	B	B	B	B	B	B	B	B

tion to the right or left side, which presents the differences between the three profiles. The MLSS concentration and SV30 were considered as alternative independent variables but were not selected because MLSS concentration and SV30 were clearly distributed according to three different settling properties, and therefore obscured the informative value of other variables.

Based on the distributions of each variable along with three categories of settling, a discriminant analysis for three groups was performed using SPSS 14.0. Two Fisher's canonical discriminant functions for non-standardized variables were derived as follows. Table 3 shows the scores and their classified groups according to the canonical discriminant functions. Only D1 was misclassified for the same reason at the dynamic time warping results.

$$Y1=-0.0344 * R\ average+-0.1931 * R\ stdev+0.0013 * R\ sum-0.5105$$
$$Y2=-0.1529 * R\ average+0.1104 * R\ stdev+0.0005 * R\ sum+3.6555$$

5. Validation of Each Feature Extraction Method

The developed algorithms by dynamic time warping and discriminant functions were validated using data prepared for validation (Table 1) and the results are shown in Table 4. In accordance with the rule to search for the smallest value among the three D values as the results of warping to the three standard patterns and its belonging group, the detection performance was presented with 69% accuracy. Conversely, when using the discriminant functions, only one case was misclassified, with an accuracy of 92%.

From the results of the classified group in Table 1, the discriminant functions have superior accuracy, whereby only one case was misclassified at the normal settling category. However, it should be noted that more representative standard patterns, including many data sets for each case and clearer case identification, could improve the accuracy. Consequently, data preparation, searching for data characteristics, and creating good standard patterns are very important and difficult tasks in properly applying the pattern matching methods to the field. In the case of discriminant analysis, the performance depends on how much the variables express the real characteristics of each group. If many variables have already been measured at each group case, the selection of variables to be used in the discriminant analysis becomes very important.

The objective of this work is not to select a method with an improved performance, but to investigate a method which can be easily applied in the field from initial works to build a detection and diagnosis system. Consequently, it was considered that discriminant analysis had a more robust performance and was an easily applicable method for settleability detection with the automated SV30 measurement.

CONCLUSION

The settling of activated sludge is a very important variable that needs to be monitored every day. When all the diagnosis work for settling depends on the monitoring of the sludge blanket height and concentration of the effluent SS, a time delay can be introduced in order to detect changes of settleability due to the specification of settling tank operational conditions. To deal with this problem, the automated SV30 measurement was developed and a settleability detection algorithm was suggested that can mimic observation work performed by the human eye and recognize the settling of property. The dynamic time warping method and the discriminant analysis were applied in order to detect settling property and their performances were compared. It should be noted that the results provided of the two methods were under limited data sets in order to provide well-fitted detection methods of the settling problems that are not properly optimized within this research. However, at the stage of starting to apply these methods, the discriminant analysis can be more easily implemented in the field due to its simplicity and small amount of test data.

ACKNOWLEDGMENT

This study was financially supported by grant from Pusan National University in program (Post-Doc. 2007) and 『I^2WaterTech of Eco-STAR project (I^2WATERTECH 04-5)』.

REFERENCES

1. T. Sekine, H. Tsugura, S. Urushibara, N. Furuya, E. Fujimoto and

Table 4. Validation results for each feature extraction method

	DTW				Discriminant analysis		
	D_st(N)	D_st(D)	D_st(B)	Group	Y1	Y2	Group
V_N1	**724.14**	2936.70	4423.50	N	3.844709	3.764732	N
V_N2	**448.57**	1156.70	4052.10	N	5.637209	1.845553	N
V_N3	**1136.10**	2325.30	3494.20	N	−0.92932	0.494467	**D**
V_D1	508.29	**495.30**	4113.40	D	6.027329	−2.49032	D
V_D2	**440.43**	541.10	4183.20	**N**	7.102059	−1.67706	D
V_D3	549.57	**504.10**	4157.90	D	7.208031	−1.19008	D
V_D4	747.43	**478.70**	3852.60	D	6.855587	−0.69712	D
V_B1	3122.10	924.90	**718.82**	B	−6.17148	−3.16014	B
V_B2	2633.40	794.30	**567.00**	B	−6.55482	−3.4968	B
V_B3	2264.00	**650.10**	930.64	**D**	−6.47941	−4.14456	B
V_B4	2237.10	**622.90**	1222.40	**D**	−6.68444	−4.52582	B
V_B5	2084.90	**590.90**	1351.80	**D**	−6.55451	−4.58964	B
V_B6	2064.70	736.30	588.45	B	−6.68506	−3.74621	B

S. Matsui, *Wat. Res*., **23**(3), 361 (1989).
2. P. Vanrolleghem, D. Van der Schueren, G. Krikilion, K. Grijspeerdt, P. Willems and W. Verstraete, *Wat. Sci. Tech*., **33**(1), 37 (1996).
3. K. Grijspeerdt and W. Verstraete, *Wat. Res*., **31**(5), 1126 (1997).
4. C. K. Yoo, S. W. Choi and I. B. Lee, *Korean J. Chem. Eng.*, **19**(3), 377 (2002).
5. M. da Motta, M. N. Pons and N. Roche, *Wat. Sci. Tech*., **46**(1-2), 363 (2002).
6. R. Jenné, E. N. Banadda, I. Y. Smets and J. F. Van Impe, *Wat. Sci. Tech*., **50**(7), 281 (2004).
7. D. P. Mesquita, A. L. Amaral, E. C. Ferreira and M. AZ. Coelho, *J. Chem. Technol. Biotechnol.*, **84**, 554 (2008).
8. D. P. Mesquita, O. Dias, A. L. Amaral and E. C. Ferreira, *Bioprocess Biosyst. Eng*., **32**, 361 (2009).
9. J. Comas, I. Rodríguez-Roda, M. Sànchez-Marrè, U. Cortés, A. Freixó, J. Arráez and M. Poch, *Wat. Res*., **37**(10), 2377 (2003).
10. S.-G. Bergh and G. Olsson, *Wat. Sci. Tech*., **33**(2), 219 (1996).
11. D. Jenkins, M. G. Richard and G. T. Daigger, *Manual on the causes and control of activated sludge bulking and foaming*, 2nd edn., Lewis Publishers, Boca Raton, USA (1993).
12. H. Sakoe and S. Chiba, *IEEE Transactions on Acoustics Speech and Signal Processing*, **26**(1), 43 (1978).
13. S. Sharma, *Applied multivariate techniques*, John Wiley & Sons, USA (1996).