Fish Observation, Detection, Recognition and Verification in The Real World

Yi-Haur Shiau[1], **Sun-In Lin**[2], **Yi-Hsuan Chen**[2], **Shi-Wei Lo**[2], **Chaur-Chin Chen**[1]

[1] Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.
No. 101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan 30013, R.O.C.
[2] National Center for High-performance Computing,
No.7, R&D Rd. VI, Hsinchu Science Park, Hsinchu 30076, Taiwan, R.O.C.
{ihow@nchc.narl.org.tw, lsi@nchc.narl.org.tw, best1013@nchc.narl.org.tw, lsw@nchc.narl.org.tw, cchen@cs.nthu.edu.tw}

Abstract - *The purpose of this paper is to present fish observation, detection, recognition and verification for processing video stream data in the real world. A distributed real-time high-definition underwater video stream system has been demonstrated in Taiwan for long-term fish observation. End users can real-time observe the high-definition underwater ecological environment via Internet. These video data is preserved to form a resource base for marine biologists. Based on the video data, fish detection is implemented. However, it is complicated in the unconstrained underwater environment, due to the water flow causes the water plants sway severely. In this paper, a bounding-surrounding boxes method is proposed to overcome the problem. It efficiently classifies moving fish as the foreground objects and the swaying water plants as the background objects. It enables to remove the irrelevant information (without fish) to reduce the massive amount of video data. Moreover, we can acquire the images of multiple species of fish with varied angles, sizes, shapes, and illumination to construct a fish category database. Sparse representation-based classification (SRC) based on compressive sensing is shown to be robust for face recognition in recent years. We propose a maximum probability of parting ranking method based on the framework of SRC for fish recognition and verification. Experimental results show that the data volume is reduced greatly, and fish recognition and verification are able to achieve high accuracy.*

Keywords: Compressive sensing, fish recognition, fish observation, sparse representation classification, real-time streaming.

1 Introduction

The research of marine ecosystems is important for understanding environmental effects, but it is difficulty due to the inaccessibility of data. In this paper, a distributed architecture for real-time high-definition underwater video stream system is demonstrated for long-term fish observation on NMMBA (National Museum of Marine Biology and Aquarium), the Southern-most coast of Taiwan [10]. Presently, real-time video streams are accessible online via Internet broadcasting. Worldwide researchers and end users can now real-time observe the underwater ecological environment. The video data is also preserved to form a resource base for marine biologists. However, the stored data that it reaches 1 gigabyte per hour is huge for storage space. In our observation environment, fish does not always appear in the video frames. Thus, fish detection is implemented that can remove the irrelevant information (without fish) to reduce the data volume.

Although many applications for object detection and tracking have been proposed, application in uncontrolled conditions, i.e. in real-life underwater systems, remains a challenge [3]. Fish detection and tracking is complicated by the variability of the underwater environment. The water plants may be regarded as foreground objects as result of the severe sway from interference of the water flow, which is able to result in the complexities and difficulties to discriminate moving fish and swaying water plants. In this paper, we propose a bounding-surrounding boxes method, which effectively achieves the purpose that classifies moving fish as the foreground objects and swaying water plants as the background objects. Then, we are able to acquire the images of multiple species of fish with varied angles, sizes, shapes, and illumination to construct a fish category database.

Compressive sensing theorem, a novel sampling technique for finding sparse solutions to underdetermined linear system, has presented in recent years [1, 2, 4, 9]. According to sparsity principle of compressive sensing, it is possible to recover certain signals and images exactly from far fewer samples of measurements beyond Nyquist rates [7]. Based on compressive sensing theorem, a sparse representation-based classification (SRC) method is proposed for robust face recognition [12, 14]. The training images are used as the dictionary of representative samples, and the testing image is coded as a sparse linear combination of the training images via l_1-norm minimization. In this paper, we propose a maximum probability of partial ranking method based on SRC algorithm for fish recognition and verification.

The rest of this paper is organized as follows: Section 2 introduces the fish observation and fish detection method. Section 3 proposes a maximum probability of partial ranking method based on SRC for fish recognition and verification. Section 4 shows experimental results and the conclusion is drawn in Section 5.

2 Fish Observation and Fish Detection

In this paper, a distributed architecture for real-time high-definition underwater video stream system is proposed for long-term fish observation. The video stream data is preserved for further implement fish detection to construction a fish category database.

2.1 Distributed Real-Time High-Definition Underwater Video Stream System

Figure 1 shows the distributed architecture components and stream pipeline. The system is composed of three units: capture devices, stream processor, and display devices. In the capture devices unit, the signals are received from high-definition cameras and they are converted to multiple video encoded formats, such as MJPEG, MPEG 1/2/4, SWF/FLV, and WMV for multiple display platforms. The stream processor unit is in charge of post-processing of video stream data. It has two options, one is directly streaming to display devices unit that uses a stream relay server to bridge the stream data between unicast and multicast. The other is to slices video stream into sequence of images for further implementing image processing methods, such as object detection, tracking, and recognition. Meanwhile, these data are converted to SWF/FLV format and are stored as historical data. Figure 2 illustrates the workflow of the stored historical data. The display devices unit supports multiple display devices handy to end users, such as web-based user interface and mobile interface.

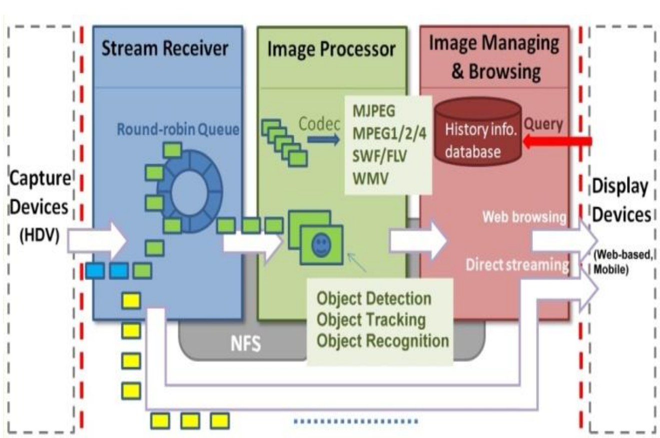


Figure 1. Architecture blocks and stream pipeline.

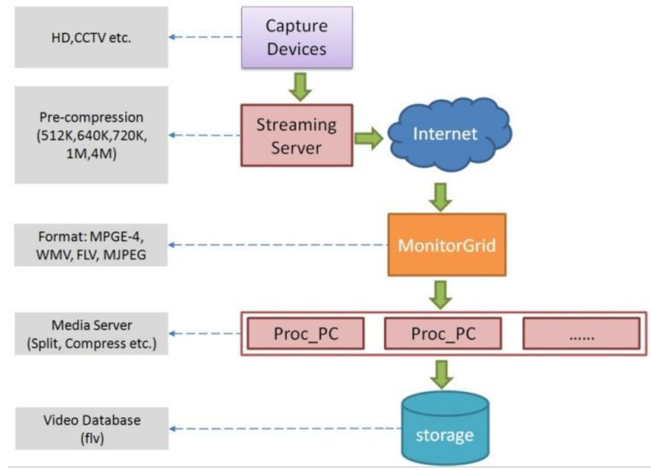


Figure 2. Web-based user interface.

Two high-definition cameras with 1280x1080 resolution located on two different sites inside a fairly large lagoon in NMMBA in Taiwan are implemented to test the above-mentioned system. Figure 3 illustrates the architecture of the underwater observation site. Two waterproof cases are set up to protect the high-definition cameras, 1394 repeaters and optical fibers. The underwater video streams are transmitted from underwater cameras to video servers on land by using optical fibers, and they are transferred back to NCHC's (National Center for High-Performance Computing) multicasting pool through ADSL lines. The marine biologists and end users can real-time observe the underwater video data via the web-based user interface or mobile interface.

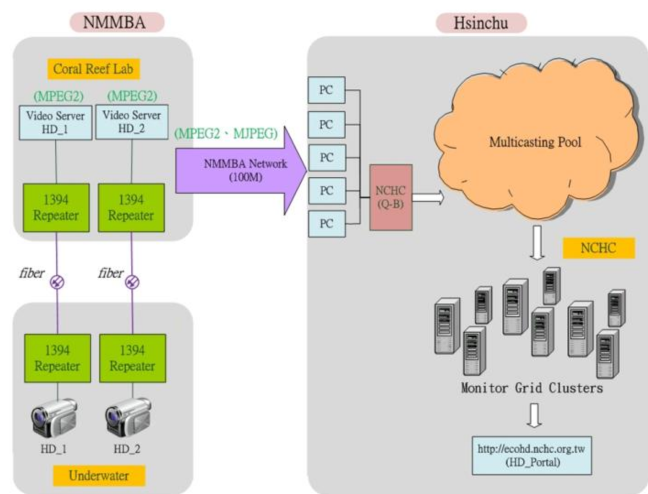
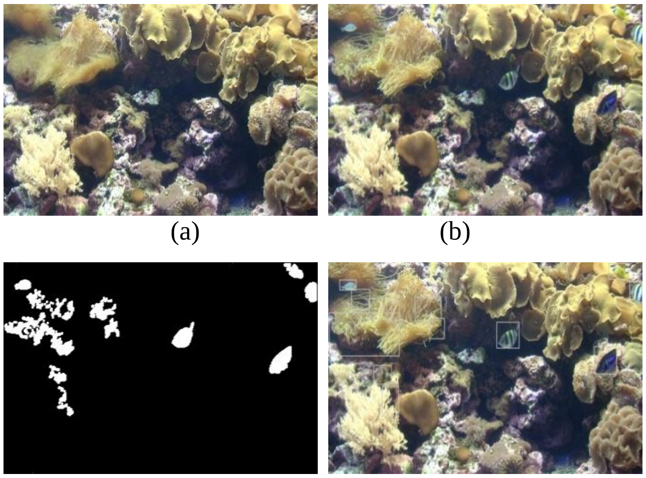


Figure 3. The architecture of the underwater observation site.

2.2 Fish Database Construction

For the stored data, background subtraction [8], foreground segmentation and object tracking methods are implemented for fish detection and tracking. In this paper, Gaussian Mixture Matrix (GMM) method is adopted for background subtraction [5, 13]. The highest color histogram similarity and the shortest distance are used for feature extraction to track the foreground objects. Figure 4(a) shows the background model and the current frame is illustrated in Figure 4(b). Figure 4(c) illustrates the foreground objects and Figure 4(d) shows the bounding boxes of these foreground objects.



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

Figure 4. (a) The background model (b) the current frame (c) the foreground objects (d) the bounding boxes of foreground objects.

The underwater environment in the real world is unconstrained, owing to the interference of the water plants sway severely. It raises the difficulty and complexity to discriminate moving fish and swaying water plants. However, the water plants always sway in a fixed field, but fish can free move to anywhere. Based on the concept, we propose a bounding-surrounding boxes method to discriminate fish as the foreground objects and water plants as the background objects. The foreground object is circumscribed by its bounding box with width w_1 and height h_1. Let (c_x, c_y) be the center point of the bounding box and the upper-left point is (c_x-0.5*w_1, c_y-0.5*h_1). Then, the surrounding box is set to T (T > 1) times the size of the bounding box with the same center point. Let B_t and S_t be the bounding box and surrounding box observed at time t. The location of S_t is fixed in the image, and the location of bounding box of the object is observed in a period of time τ. If the location of the bounding box from time t to time t+τ is always inside the range of S_t, the object is classified as a non-fish object (water plants). It is not only identified as a background object, but also eliminated from the tracked object. On the other hand, if the location of the bounding box has left the range of S_t, the object is classified as a foreground object (fish). The detecting results are shown in Figure 5. The yellow box represents the fixed surrounding box of the object. The red box in Figure 5(a) represents the object is classified as “fish”, and the blue box in Figure 5(b) represents the objects is classified as “non-fish” object.

(a) (b)

Figure 5. (a) The object (red box) is classified as fish (b) the object (blue box) is classified as non-fish (water plant).

After fish detection is implemented using our proposed method, we can only record the video data that contains fish and remove the irrelevant information (without fish) to reduce the stored data volume. We acquire the images of multiple species of fish with varied angles, sizes, shapes, and illumination. For each species of fish, we select some images that are almost different, and implement image resizing method to resize all of fish images to the same resolution. Then, a fish category database in the real world is constructed.

3 Maximum Probability of Partial Ranking Method

An SRC method that represents a testing image as a sparse linear combination of all training images has been shown to be robust for face recognition [7, 12, 14]. In this paper, we proposed a maximum probability of partial ranking method based on SRC method for fish recognition. There are K species of fish in our fish category database. Therefore, we set $B = [B_1, B_2, \dots, B_K]$ as the concatenation of the N training images from K species of fish, where $N = n_1+n_2+\dots+n_K$. The training images of the i^{th} species of fish is defined as $B_i = \left[s_1^{(i)}, s_2^{(i)}, \dots, s_{n_i}^{(i)}\right] \in \boldsymbol{R}^{m\times n_i}$. $s_j^{(i)}$ is an m-dimensional vector stretched by the j^{th} image of the i^{th} species of fish. A testing image $\boldsymbol{y} \epsilon \boldsymbol{R}^m$ of the i^{th} species of fish could be represented as a linear combination of the training images in B_i, i.e. $\boldsymbol{y} = \sum_{j=1}^{n_i} \alpha_j^{(i)} s_j^{(i)} = B_i\, \boldsymbol{\alpha}^{(i)}$, where $\boldsymbol{\alpha}^{(i)} = [\alpha_1^{(i)}, \alpha_2^{(i)}, \dots, \alpha_{n_i}^{(i)}]^T \in \boldsymbol{R}^{n_i}$ are weight coefficients. Let $\boldsymbol{y} = B\boldsymbol{\alpha}$ represent the testing image $\boldsymbol{y}$ by using B, where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}^{(1)}; \boldsymbol{\alpha}^{(2)};\dots;\boldsymbol{\alpha}^{(K)}]$. Due to $\boldsymbol{y}$ belongs to the i^{th} species of fish and $\boldsymbol{y} = B_i\boldsymbol{\alpha}^{(i)}$, a perfect solution to $\boldsymbol{\alpha}$ is that only the coefficients in $\boldsymbol{\alpha}^{(i)}$ have significant values, and all the coefficients in $\boldsymbol{\alpha}^{(j)}$, j=1,2,…,K and $j \neq i$, are nearly zero.

An SRC method computes the residuals as a classifier to accurately assign $\boldsymbol{y}$ to the certain species of fish. In this paper, we compute maximum probability of partial ranking to replace the residuals as a classifier (called SRC-MP). It is found by experiments that the largest coefficient may not belong to the exact species of fish. However, the γ largest coefficients may almost match the correct species of fish. Thus, we convert and normalize the coefficient $v_j^{(i)}$ into the probability value $p_j^{(i)} = \frac{v_j^{(i)}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} v_j^{(i)}}$, where $v_j^{(i)}$ is the j^{th} non-zero coefficient greater than zero of the i^{th} species of fish of $\hat{\boldsymbol{\alpha}}_1$. Then, we assign a partial ranking value γ (first largest values), and sum up these largest γ values to obtain a new probability value for each species of fish, respectively. Moreover, we employed the new maximum probability as the classifier.

The complete method we proposed is summarized as bellow,

1. Set $B_i = \left[s_1^{(i)}, s_2^{(i)}, \ldots, s_{n_i}^{(i)}\right] \in \boldsymbol{R}^{m \times n_i}$ as a matrix of the training images for *K* species of fish, and a testing image $\boldsymbol{y} \in \boldsymbol{R}^m$, as input data.

2. Solve the l_1-norm minimization problem.

   $$\hat{\boldsymbol{\alpha}}_1 = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \|B\boldsymbol{\alpha} - \boldsymbol{y}\|_2 \leq \varepsilon. \quad (1)$$

3. Compute the probability value $p_j^{(i)} = \frac{v_j^{(i)}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} v_j^{(i)}}$ for all non-zero values greater than zero.

4. Compute new probability value for each species of fish $w_i(\boldsymbol{y})$, respectively.

   for $k \leq \gamma$, $w_i(\boldsymbol{y}) = w_i(\boldsymbol{y}) + p_k^{(i)}$ for $i = 1, \ldots, K$, where $p_k^{(i)}$ is the k^{th} largest probability value belonging to the i^{th} species of fish.

5. Label $\boldsymbol{y}$ by identity($\boldsymbol{y}$) = arg{$\max_i w_i(\boldsymbol{y})$}.

4 Experimental Results

A bounding-surrounding boxes method for fish detection is implemented on the video data. It efficiently discriminates moving fish as foreground objects and swaying water plants as background objects. It enables to remove the irrelevant information (without fish) in the video data to reduce the data volume. The experimental result shows that we can reduce the data volume to about 1/10 averagely. Furthermore, we acquire the images of multiple species of fish with varied angles, sizes, shapes and illumination to construct a fish category database in the real world. Based on the database, a maximum probability of parting ranking method based on SRC method is implemented for fish recognition and verification.

4.1 Fish Category Database

Prior to fish recognition, sufficient data for constructing a database of fish category is necessary. The fish category database that we constructed is composed of 1,000 fish images of 180 rows and 130 columns with JPEG file format. Totally, there are 25 different species of fish. Each one contributed 40 images with varied angles, sizes, shapes and illumination. The 5 training fish images of subjects 2, 10, 11, 14, 19, 24 are illustrated in Figure 6. The total 40 fish images of subject 2 are illustrated in Figure 7.

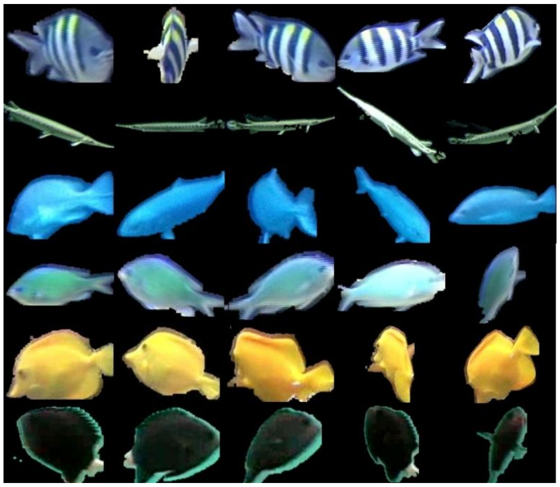


Figure 6. The 5 training fish images of 6 subjects.

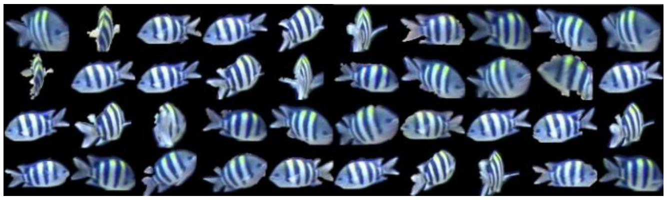


Figure 7. Examples of total 40 fish images of subject 2.

4.2 Fish Recognition

We evaluate the performance of our proposed method (SRC-MP) for fish recognition on the fish category database. For each species of fish, we randomly selected 20 images for training, while the rest 20 images for testing. Eigenfaces [11] and fisherfaces [6] are used for feature extraction with the feature space dimensions d = 12, 16, 20, 30, 40, 50, respectively. We assign the partial ranking value γ = 10 to compute the recognition rate. Table 1 shows the recognition rates of all methods: (1) Eigen + SRC-LV, (2) Eigen + SRC-MP, (3) Fisher + SRC-LV and (4) Fisher + SRC-MP. verse the corresponding feature dimensions. Figure 8 shows the curve of the recognition rates, and the maximum recognition rate enables to approach over 80%.

Table 1. Recognition rates (%) of all methods on the fish category database associated with the corresponding dimensionality.

	d = 12	d = 16	d = 20	d = 30	d = 40	d = 50
(1)	61.6	71.0	73.2	77.0	77.2	80.0
(2)	63.2	73.6	75.8	79.2	80.4	81.6
(3)	58.2	60.2	63.0	68.2	77.4	79.6
(4)	58.6	61.8	66.0	72.8	79.0	81.8

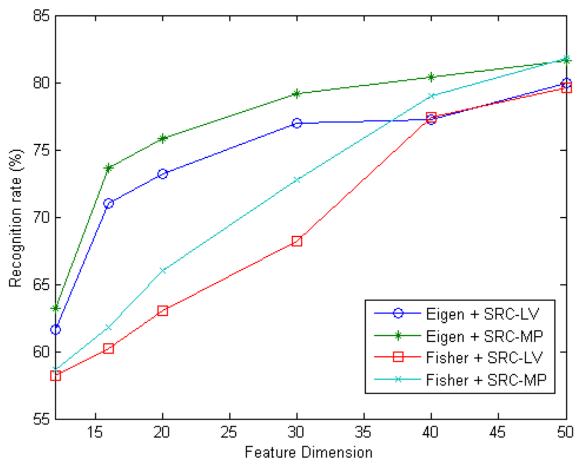


Figure 8. Recognition rates of all methods versus feature dimension on the fish category database.

4.3 Fish Verification

Fish verification, verify whether the testing image is a valid image of one of the species of fish in the database, is a problem that different from identification. As for each species of fish on the database, the first 30 images for training and the next 10 images (valid images) for testing were selected. We also collected new 13 species of fish with 10 images (invalid images) that are not species of fish on the database for testing. Eigenfaces and fisherfaces are used for feature extraction with the feature dimensional d = 12, 16, 20, 30, 40. Figure 9 shows the weighting coefficient of an invalid testing image. The weighting coefficients are not concentrated on any one subject and instead spread widely across the entire training set. Thus, the nonzero weighting coefficients of a valid testing image concentrate mostly on one species of fish, whereas an invalid image has weighting coefficients widely spread among multiple species of fish.

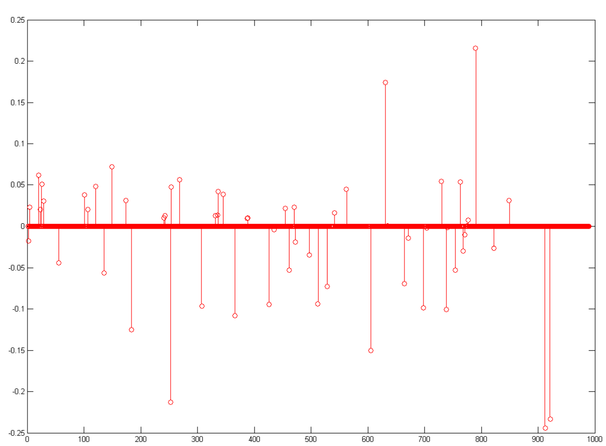


Figure 9. Weighting coefficients of an invalid testing image.

First, we obtain the value $c_j^{(i)}$ that the j^{th} image of the i^{th} testing fish belongs to one species of fish on the database, where j = 1, 2, …, n_i. The value of $c_j^{(i)}$ is k, if $c_j^{(i)}$ belongs to the k^{th} species of fish on the database. For valid fish, we compute the number $C^{(i)}$ that the i^{th} valid testing fish is classified to the correct species of fish. For invalid fish, $C^{(i)}$ is the maximum number of the i^{th} invalid testing fish belongs to one species of fish. Then, we compute the probability value of the i^{th} testing fish is ${C^{(i)}}/{n_i}$. Table 4 shows the probability value (italic value) of all valid and invalid testing fish, that eigenfaces is used for feature extraction with the feature dimensional d = 30. Two mean values $\sum_{i=1}^{z} C^{(i)}/_z$, valid mean value and invalid mean value, are computed, where z is the number of the valid testing fish and invalid testing fish, respectively. A threshold t that is the mean value of the valid mean value and the invalid mean value is implemented. For valid testing fish, if the probability value of $C^{(i)}$ is higher than t, the valid testing fish is represented as a valid fish, otherwise, it is represented as an invalid fish. For invalid fish, if the probability value of $C^{(i)}$ is lower than t, the invalid testing fish is represented as an invalid fish, otherwise, it is represented as a valid fish. In table 2, the valid mean value is 90 and the invalid mean value is 43.85. The threshold t is (85.6 + 39.2)/2 = 62.4 is assigned, and the verification rate that the valid fish is represented as valid fish is 92% (23/25), and the verification rate that invalid fish is represented invalid fish is 92%. Thus, we can efficiently verify the testing image is valid or invalid on the database. Table 3 shows the verification rates of all methods: (1) Eigen (Valid), (2) Eigen (Invalid), (3) Fisher (Valid) and (4) Fisher (Invalid).

Table 2. The Probability values (italic values) of all valid and invalid subjects. The bold values are the testing subjects that they are lower than threshold t for valid testing subjects, and they are higher than t for invalid testing subjects.

Valid individual											
id	%	id	%	id	%	id	%	id	%	id	%
1	***60***	2	*80*	3	*80*	4	*100*	5	*100*	6	*100*
7	*100*	8	*100*	9	*90*	10	*100*	11	*80*	12	*80*
13	*90*	14	*80*	15	*70*	16	*80*	17	*70*	18	*90*
19	***50***	20	*90*	21	*90*	22	*90*	23	*80*	24	*100*
25	*90*										
Invalid individual											
id	%	id	%	id	%	id	%	id	%	id	%
1	*40*	2	*60*	3	*50*	4	*10*	5	*30*	6	*30*
7	*40*	8	***100***	9	***70***	10	*50*	11	*30*	12	*30*
13	*30*	14	*20*	15	*50*	16	*40*	17	*30*	18	*30*
19	*20*	20	*50*	21	*20*	22	*30*	23	*40*	24	*60*
25	*20*										

Table 2. The verification rates of all methods: (1) Eigen (Valid), (2) Eigen (Invalid), (3) Fisher (Valid) and (4) Fisher (Invalid).

	d = 12	d = 16	d = 20	d = 30	d = 40	d = 50
(1)	66.0	80.0	80.0	92.0	88.0	96.0
(2)	92.0	88.0	80.0	92.0	88.0	92.0
(3)	56.0	68.0	72.0	76.0	84.0	88.0
(4)	72.0	76.0	76.0	88.0	88.0	88.0

5 Conclusion

In this paper, a distributed real-time high-definition underwater video stream system was developed for long-term fish observation in the real world. The stored video data was huge for storage space. A bounding-surrounding boxes method had been proposed to discriminate moving fish as the foreground objects and swaying water plants as the background objects. Then, it enabled to efficiently remove the irrelevant information (without fish) and only save the data containing fish. It reduced the massive amount of the video data greatly. After that, we acquired the images of multiple species of fish with varied angles, sizes, shapes, and illumination to construct a fish category database. We presented a maximum probability of partial ranking method based on the framework of sparse representation-based classification (SRC) for fish recognition and verification. Eigenfaces and fisherfaces were utilized for feature extraction. Experimental result showed our proposed method achieved high recognition and verification accuracy. In the future work, we plan to identify fish species real-time from the live video data.

6 Acknowledgments

We thank National Museum of Marine Biology & Aquarium, Taiwan and Ecogrid team at National Center for High-Performance Computing, Taiwan for fish image sequences data supply.

7 References

[1] S. Auethavekiat, "Introduction to the Implementation of Compressive Sensing," *AU Journal of Technology*, pp. 39-46, 2010.

[2] E.J. Candes and M.B. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, vol.5, no.2, pp. 21-30, 2008.

[3] D.R. Edgington, I. Kerkez, D.E. Cline, J. Mariette, M. Ranzato, and P. Perona, "Detecting, tracking and classifying animals in underwater video," *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 634-638, 2007.

[4] S.J. Kim, K. Koh, M. Lustig, and S. Boyd, "An efficient method for compressed sensing," *IEEE International Conference on Image Processing*, vol.3, pp. 117-120, 2007.

[5] D.S. Lee, J.J. Hull, and B. Erol, "A Bayesian framework for gaussian mixture background modeling," *IEEE International Conference on Image Processing*, vol.3, pp. 973-976, 2003.

[6] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.R. Mullers, "Fisher Discriminant Analysis with Kernels," *IEEE International Workshop on Neural Networks for Signal Processing*, vol.9, pp. 41-48, 1999.

[7] C. Moler, ""Magic" reconstruction: compressive sensing," *Cleves Corner, Mathworks News&Notes*, pp. 1-4, 2010, http://www.mathworks.com

[8] M. Piccardi, "Background subtraction techniques: a review," *IEEE International Conference on System, Main, and Cybernetices*, pp. 3099-3104, 2004.

[9] O. Samet, K. Amin, and H. Babak, "Weighted Compressed Sensing and Rank Minimization," *International Conference on Acoustics, Signal and Speech Processing*, pp. 3736-3739, 2011.

[10] Y.H. Shiau, J.S. Cheng, S.I. Lin, Y.H. Chen, K.T. Tseng, H.M. Chou, and S.W. Lo, "A Distributed Architecture for Real-Time High-Resolution Video Streaming," *International Conference on Parallel & Distributed Processing Techniques & Applications*, pp. 345-349, 2009.

[11] M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, vol.3, no.1, pp. 71-86, 1991.

[12] J. Wright, A.Y. Yang, and A. Ganesh, "Robust Face Recognition via Sparse Representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.31, no.2, pp. 210-227, 2009.

[13] S.Y. Yang and C.T. Hsu, "Background Modeling from GMM Likelihood Combined with Spatial and Color Coherecny," *IEEE International Conference on Image Processing*, pp. 2801-2804, 2006.

[14] Q. Zhang and B. Li, "Joint Sparsity Model with Matrix Completion for an Ensemble of Face Images," *IEEE International Conference on Image Processing*, pp. 1665-1668, 2010.