Using Sparse Representation for Fish Recognition and Verification in Real World Observation

Yi-Haur Shiau[1,2], Fang-Pang Lin[2] and Chaur-Chin Chen[3]

[1]Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan
[2]Data Computing Division, National Center for High-Performance Computing, Hsinchu, Taiwan
[3]Institute of Information Systems and applications, National Tsing Hua University, Hsinchu, Taiwan

Abstract - *The purpose of this paper is to present an innovated fish recognition and verification method suited for the real world automatic underwater fish observation. Based on the fish recognition and verification, biologists can study fish population as well as identify new species of fish appearing in area of interest. A distributed real-time underwater video stream system has been developed in Taiwan for long-term ecological observation. The system also archives video data and incorporates data analysis. We propose a fish detection procedure on the video data to obtain multiple species of fish images with varied angles, sizes, shapes, and illumination, which leads to a fish category database. In recent years, a sparse representation-based classification (SRC) based on compressive sensing is developed. Based on the SRC method, we propose a maximum probability of parting ranking method for fish recognition and verification, in which the eigenfaces and fisherfaces are used for feature extraction on the category database. Experimental results show that the proposed fish recognition and verification method is able to achieve high accuracy and robustness.*

Keywords: Compressive sensing, sparse representation classification, fish recognition, fish verification.

1 Introduction

Ecological observation is imperative for marine scientists to study marine ecosystems. It is however difficult to sustain a long term and real-time observation, mostly due to inaccessibility of the marine environment [1]. In the past few years, a distributed real-time underwater video stream system has been developed for long-term observation of ecosystem at the Southern tropical coast of Taiwan [2]. Not only can the video data be broadcasted in real-time via the Internet, but also archived to form a resource base for further analysis. A fish detection method that is able to efficiently discriminate moving fish and drift water plants is incorporated with the system to obtain multiple species of fish images with varied angles, shapes, and illumination. The detected images are subsequently collected to construct a fish category database on the fly.

In recent years, a novel sampling method, compressive sensing, is proposed to find sparse solutions of underdetermined linear systems and proved to be effective [3]. The compressive sensing is based on the assumption of sparsity of signals of interest and reconstructs the signals exactly from far fewer samples of measurements below the Nyquist rate [4]. A sparse representation-based classification (SRC) method based on compressive sensing is proposed and successfully applied in face recognition [5], in which 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. Following the SRC method, we propose a maximum probability of partial ranking method for fish recognition and verification based on the fish category database.

The rest of the paper is organized as follows: Section 2 briefly introduces fish detection method on distributed real-time underwater video stream system. Section 3 describes the 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 Detection on Underwater Video Stream System

In this paper, the fish category database is constructed through a distributed real-time underwater video stream system, which enables a long-term underwater fish observation. Figure 1 shows the architecture of the video stream system. The video signals captured from CCTVs are converted into Motion JPEG stream. However, the stream size is huge and it has interlace effect that affects fish detection and tracking. We transfer the stream into multiple encoded formats and bitrates as well as implement inter-field deinterlacing method [6] to remove interlace effect. We compute Peak Signal-to-Noise Ratio (PSNR) between native Motion JPEG and other encoded formats to obtain a suitable solution, better quality and lower data size, for our system. Figure 2 shows the PSNR with different encoded formats and bitrates (left) and deinterlacing result (right up: original, right down: deinterlacing). In our system, mpeg4 format with 5Mb bitrate (PSNR=31.87) enables to achieve our requirement, real-time observation and further fish analysis. Then, there are two processing mode for these mpeg4 streams, one is converted into mpeg4 video files for storing into local storage, the other is directly transmitted into multicasting pool via ADSL lines for real-time observation.

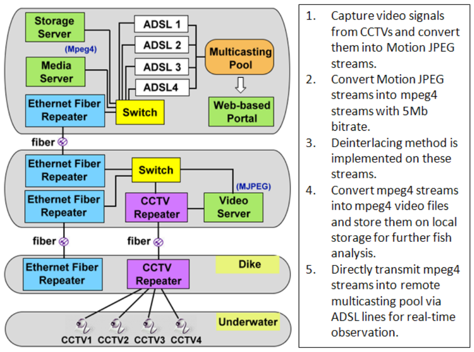

Figure 1. Architecture of the underwater video stream system.

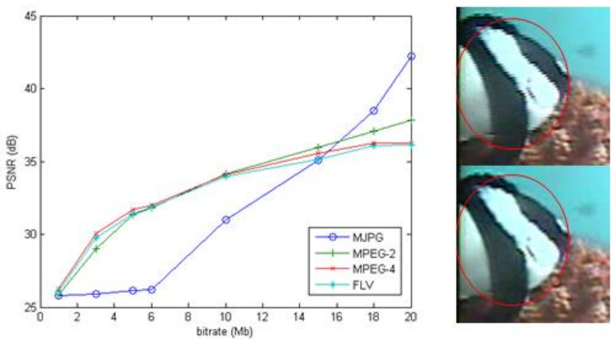

Figure 2. The PSNR (left) and deinterlacing (right) results.

For the stored video data, fish detection is implemented. Gaussian Mixture Matrix (GMM) method [7] is utilized for background subtraction. Each pixel is modeled by a mixture of G Gaussian distributions. The history of a pixel is defined as a time series $\{X_1, \ldots, X_t\}$. The probability function of the pixel value in frame t is

$$p(X_t) = \Sigma_{i=1}^{G} w_t^i * \eta(X_t, \mu_t^i, \Sigma_t^i) \quad (1)$$

Where w_t^i is the weight, μ_t^i is the mean value, Σ_t^i is the covariance matrix of i^{th} Gaussian distribution at time t, and η is a Gaussian probability density function. Figure 3(a) shows the background model and Figure 3(b) illustrates the detected foreground objects that include fish and water plants.

The underwater environment in the real world is unconstrained, owing to the interference of the water plants drift severely. It raises the difficulty and complexity to discriminate moving fish and drift water plants. We proposed a bounding-surrounding boxes method that based on the concept that water plants always drift in a fixed field, but fish can move anywhere. Each foreground object is circumscribed by its bounding box with width d_1 and height h_1. Let (o_x, o_y) be the center point of the bounding box and the upper-left point is $(o_x\text{-}0.5{*}d_1,\ o_y\text{-}0.5{*}h_1)$. Then, the surrounding box is set to T ($T > 1$) times the size of the bounding box with the same center point. Let B and S be the bounding box and surrounding box. The size and location of S is fixed in the image. The location of B of the object is observed in a period of time τ. If the location of the B is always inside the range of S, the object is classified as a non-fish object (water plants) and it is eliminated from the foreground object. On the other hand, if the location of the B has left the range of S, the object is classified as a foreground object (fish). Figure 3(c) and 3(d) show the detecting results after tracking during a period of time τ. The yellow box represents the fixed surrounding box of the object. The red box and red line in Figure 3(c) represent the bounding box and the trajectory of foreground object, and the object is classified as "fish". The blue box and blue line in Figure 3(d) represent the bounding box and the trajectory of foreground object, and the object is classified as "non-fish" object. Then, mean shift method is implemented for fish tracking [8].

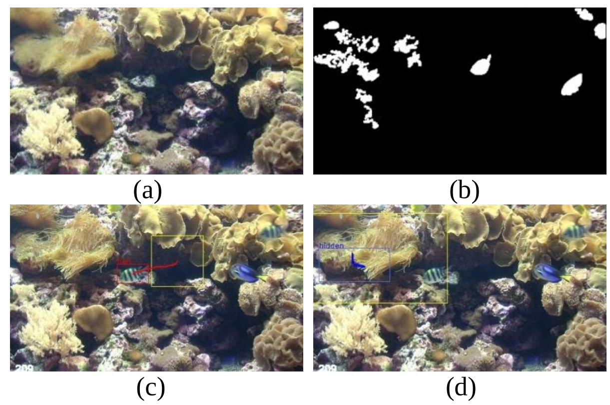

Figure 3. (a) The background model (b) the foreground objects (c) The object (red box) is classified as fish (d) the object (blue box) is classified as non-fish (drift water plant).

3 Fish Recognition and Verification

In this paper, a maximum probability of partial ranking method based on SRC is proposed for fish recognition and verification. A SRC method represents a testing image as a sparse linear combination of all training images and obtains a sparse solution via l_1-norm minimization [5]. We could represent our fish images the same as SRC proposed method, and implement for fish recognition and verification. Suppose there are K species of fish in the fish category database, and we set $A = [A_1, A_2, \ldots, A_K]$ as the concatenation of the N training images from K species of fish, where $N = n_1+n_2+\ldots+n_K$. The training images of the i^{th} species of fish is defined as $A_i = [\boldsymbol{s}_1^{(i)}, \boldsymbol{s}_2^{(i)}, \ldots, \boldsymbol{s}_{n_i}^{(i)}] \in \boldsymbol{R}^{m \times n_i}$. $\boldsymbol{s}_j^{(i)}$ is an m-dimensional vector stretched by the j^{th} image of the i^{th} species of fish. A testing image $\boldsymbol{y} \in \boldsymbol{R}^m$ of the i^{th} species of fish could be represented as a linear combination of the training images in A_i, i.e. $\boldsymbol{y} = \sum_{j=1}^{n_i} \alpha_j^{(i)} \boldsymbol{s}_j^{(i)} = A_i\, \boldsymbol{\alpha}^{(i)}$, where $\boldsymbol{\alpha}^{(i)} = [\alpha_1^{(i)}, \alpha_2^{(i)}, \ldots, \alpha_{n_i}^{(i)}]^T \in \boldsymbol{R}^{n_i}$ are weight coefficients. Let $\boldsymbol{y} = A\boldsymbol{\alpha}$ represents the testing image $\boldsymbol{y}$ by using A, where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}^{(1)}; \boldsymbol{\alpha}^{(2)}; \ldots; \boldsymbol{\alpha}^{(K)}]$. In noiseless case, $\boldsymbol{y}$ belongs to the i^{th} species of fish that only the coefficients in $\boldsymbol{\alpha}^{(i)}$ have significant values, and all the coefficients in $\boldsymbol{\alpha}^{(j)}$, $j=1,2,\ldots,K$ and $j \neq i$, are nearly zero.

In noise case, the coefficients might distribute on multiple species of fish and it is difficult to accurately assign $\boldsymbol{y}$ to correct species of fish. Figure 4 shows the coefficients distribution of $\boldsymbol{y}$ in noise case. The testing image $\boldsymbol{y}$ belongs to, for example, the 19th species of fish on the database, but it is assigned to 7th species of fish if only the largest weighting coefficient is adopted, called SRC-LV. Due to the first largest coefficients might almost distribute on the correct species of fish, a maximum probability of partial ranking method as a classifier, called SRC-MP, is proposed to overcome the situation. First, the probability value $p_j^{(i)} = \frac{c_j^{(i)}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} c_j^{(i)}}$ is computed, where $c_j^{(i)}$ 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. The new maximum probability is adopted as the classifier for fish recognition.

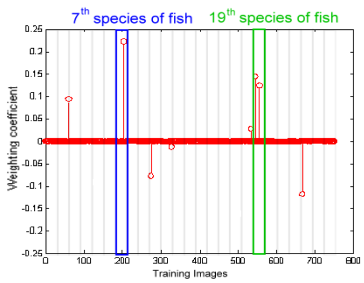

Figure 4. The weighting coefficient distribution of the testing image $\boldsymbol{y}$ in noise case.

Fish verification verifies whether the testing image is one of the species of fish on the database (valid verification) or is a new species of fish (invalid verification). The weighting coefficients of an invalid testing image are not concentrated on specific species of fish, but instead distributed widely across the training images. For fish verification, we compute two verification rates: valid verification rate (VVR) and invalid verification rate (IVR). VVR represents the testing image is assigned to correct species of fish on the database, and IVR represents the testing image is reject due to it is a new species of fish. The complete fish recognition and verification method we proposed is summarized as bellow,

1. Set $A_i = \left[\boldsymbol{s}_1^{(i)}, \boldsymbol{s}_2^{(i)}, \dots, \boldsymbol{s}_{n_i}^{(i)}\right] \in \boldsymbol{R}^{m \times n_i}$ 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.

   $$\hat{\boldsymbol{\alpha}}_1 = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \|A\boldsymbol{\alpha} - \boldsymbol{y}\|_2 \leq \varepsilon. \qquad (2)$$

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

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

   for $k <= \gamma$, $w_i(\boldsymbol{y}) = w_i(\boldsymbol{y}) + p_k^{(i)}$ for $i = 1, \dots, K$, 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{$\max_i w_i(\boldsymbol{y})$}.

The verification rates are computed as follows,

6. Compute valid total number $v_u^{(i)}$ and invalid maximum number $v_e^{(i)}$.

   Where $v_u^{(i)}$ is the total number of the i^{th} valid testing species of fish that is classified to the i^{th} species of fish on the database, and $v_e^{(i)}$ is the maximum number that the i^{th} invalid testing species of fish is classified to a certain species of fish.

7. Compute the valid probability value $q_u^{(i)} = \frac{v_u^{(i)}}{n_i}$ and invalid probability value $q_e^{(i)} = \frac{v_e^{(i)}}{n_i}$.

8. Compute the valid mean value $\rho_u = \frac{\sum_{i=1}^{z_u} q_u^{(i)}}{z_u}$ and invalid mean value $\rho_e = \frac{\sum_{i=1}^{z_e} q_e^{(i)}}{z_e}$, where z_u and z_e are the number of valid and invalid testing species of fish, respectively.

9. Assign a threshold value $= \frac{(\rho_u + \rho_e)}{2}$.

10. Compute VVR and IVR.

    VVR: VVR$= \frac{r_u}{z_u}$, where r_u is the number that $q_u^{(i)} \geq \vartheta$.

    IVR: IVR$= \frac{r_e}{z_e}$, where r_e is the number that $q_e^{(i)} \leq \vartheta$.

4 Experimental Results

The fish category database that we constructed is composed of 1,000 fish images. Each image consists of 180 rows and 130 columns pixels recorded in JPEG file format. Totally, there are 25 different species of fish. Each species was given 40 images with varied angles, shapes and illumination. The total 40 fish images of subject 2 are illustrated in Figure 5 as an example.

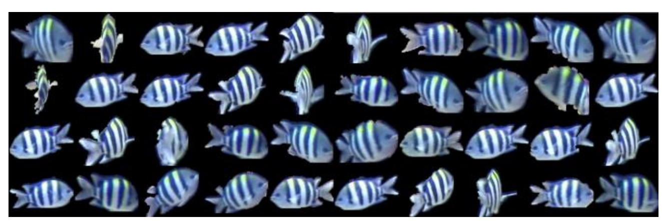

Figure 5. An example of total 40 fish images of subject 2.

4.1 Fish Recognition

We evaluate the performance of SRC-MP method 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 [9] and fisherfaces [10] 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. Figure 6 shows the curve of the recognition rates of all methods.

Table 1. Recognition rates (%) of all methods 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

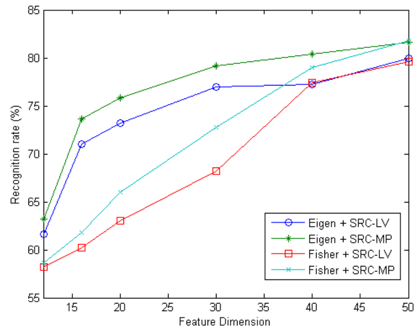

Figure 6. Recognition rates of all methods versus feature dimension on the fish category database.

4.2 Fish Verification

For each species of fish on the database, the first 30 images for training and the next 10 images (i.e. valid images by default) for testing were sequentially selected in practice. We additionally collected 25 species of fish, different from those on the database, with 10 images each (i.e. prescribed invalid images) for the invalid verification testing. Table 2 shows verification rates of all methods: (1) Eigen (Valid), (2) Eigen (Invalid), (3) Fisher (Valid) and (4) Fisher (Invalid). Figure 7 shows the curve of verification rates of all methods.

Table 2. Valid and invalid verification rates (%) of all methods associated with the corresponding dimensionality.

	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

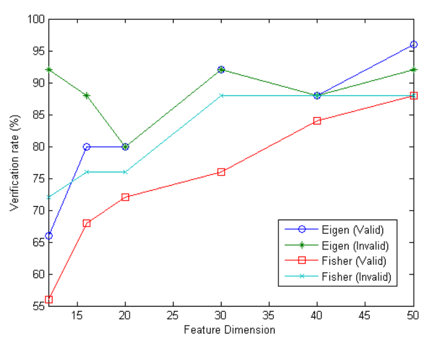

Figure 7. The verification rates of all methods versus feature dimension on the fish category database.

5 Conclusion

In this paper, a maximum probability of partial ranking method based on sparse representation-based classification (SRC) was proposed for fish recognition and verification. The method was implemented on the real world fish category database. The database was constructed by implementing a fish detection procedure using bounding-surrounding boxes method on the video data that was acquired from a distributed real-time underwater video stream system in Taiwan. Experimental results showed the proposed method was able to achieve high recognition and verification accuracy and robustness. In the future work, we plan to identify fish species in real-time from the live video data.

6 Acknowledgments

The research was funded by the Taiwan National Science Council (grant NSC 100-2933-I-492-001) and the European Commission (FP7 grant 257024) and undertaken in the Fish4Knowledge project (www.fish4knowledge.eu).

7 References

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

[2] Shiau, Y. H., Chen, Y. H., Tseng, K. T., Cheng, J. S., Lin, S. I., Lo, S. W., & Chou, H. M. (2010). A Real-Time High-Resolution Underwater Ecological Observation Streaming System. *International Society for Photogrammetry and Remote Sensing*, 517-521.

[3] Candes, E. J., & Wakin, M. B. (2008). An Introduction to Compressive Sampling. *IEEE Signal Processing Magazine*, 5, 2, 21-30.

[4] Moler, C. (2010). ”Magic” reconstruction: compressive sensing. *Cleves Corner, Mathworks News&Notes*, 1-4.

[5] Wright, J., Yang, A. Y., & Ganesh, A. (2009). Robust Face Recognition via Sparse Representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31, 2, 210-227.

[6] Lin, S. F., Chang. Y. L. & Chen L. G. (2003). Motion Adaptive Interpolation with Horizontal Motion Detection for Deinterlacing. *IEEE Trans. On Computer Electronics*, 49, 4, 1256-1265.

[7] Stauffer, C., & Grimson, W. (1999). Adaptive Background Mixture Models for Real-Time Tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 246-252.

[8] Cheng, Y. (1995). Mean Shift, Mode Seeking, and Clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17, 8, 790-799.

[9] Turk, M., & Pentland, A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3, 1, 71-86.

[10] Mika, S., Ratsch, G., Weston, J., Scholkopf, B., & Mullers, K.R. (1999). Fisher Discriminant Analysis with Kernels. *IEEE International Workshop on Neural Networks for Signal Processing*, 9, 41-48.