A SPARSE SAMPLE COLLECTION AND REPRESENTATION METHOD USING RE-WEIGHTING AND DYNAMICALLY UPDATING OMP FOR FISH TRACKING

Yi-Hao Hsiao and Chaur-Chin Chen

Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan

ABSTRACT

Tracking fish in their natural environment is an important aspect of marine ecosystem research. However, real-world fish tracking is challenging due to unconstrained environments and complex scenarios. The purpose of this study is to develop a sparse sample collection and representation method (SSCR) based on the compressive sensing concept for fish tracking. The SSCR consists of sample collection and sparse sample representation procedures. The sample collection procedure obtains sets of positive, negative, and predictive samples by using the proposed speed-up background modeling method (SuBM). The SuBM adopts nonparametric histogram concept for each pixel to build a background model, and efficiently accelerates the tracking speed. In addition, the sparse sample representation procedure represents each predictive sample as a sparse linear combination of all positive and negative samples. The weights of the predictive samples are computed using our proposed re-weighting and dynamically updating orthogonal matching pursuit method (RwDuOMP). The RwDuOMP, which includes three concepts (picking extra samples, re-weighting the picked samples, and dynamically updating negative samples), efficiently improves the performance of sparse signal reconstruction. The predictive sample with the maximum weight is regarded as the target object tracking result. We evaluate the SSCR method using several complicated real-world underwater sequences. Furthermore, we compare the SuBM with the Gaussian Mixture Model, and also compare the RwDuOMP method with the orthogonal matching pursuit (OMP), regularized OMP, and compressive sampling matching pursuit methods. Experimental results indicate that our proposed method achieves efficiently higher tracking results than other methods, and accelerates fish tracking.

Index Terms**—** Compressive sensing, object tracking, orthogonal matching pursuit, Gaussian mixture model, background subtraction

1. INTRODUCTION

Object tracking is an important issue in computer vision applications. However, under uncontrolled conditions, i.e., in a real-world underwater environment, there are many challenges. Specifically, the implementation of fish tracking is challenging due to factors such as drastic fish shape variation, fast-moving fish, and the presence of similar objects or background scenes. In recent years, numerous outstanding algorithms [1] for object tracking have been proposed. Background subtraction is regarded as one approach that can capture the complete shape of objects being tracked [2]. Particularly, the Gaussian mixture model (GMM) [3] is proposed to build a dynamic background model. In addition, compressive sensing (CS) [4], a novel sampling method, has been widely used in many fields. Based on CS theory, a real-time compressive tracking method (CT) [5] and a sparse representation-based classification method (SRC) [6] are proposed. The CT method can overcome the drift problem and achieve real-time tracking results, while the SRC method has been shown to be robust for face recognition.

In this study, a sparse sample collection and representation method (SSCR) based on CT and SRC is proposed for object tracking. The SSCR method includes two procedures: sample collection and sparse sample representation. In the sample collection procedure, a speed-up background modeling method (SuBM), which adopts nonparametric histogram concept, is proposed to accelerate fish tracking. In the sparse sample representation procedure, a re-weighting and dynamically updating orthogonal matching pursuit method (RwDuOMP) is proposed to improve the performance of sparse signal reconstruction. Figure 1 shows the main components of our proposed method. Several challenging underwater sequences, which are obtained from an uncontrolled open sea in Taiwan [7], are used to evaluate the performance of SSCR. We also compare SuBM with GMM and compare RwDuOMP with orthogonal matching pursuit (OMP) [8], regularized OMP (ROMP) [9], and compressive sampling matching pursuit (CoSaMP) [10]. Experimental results indicate that the SSCR effectively improves the accuracy and acceleration of fish tracking.

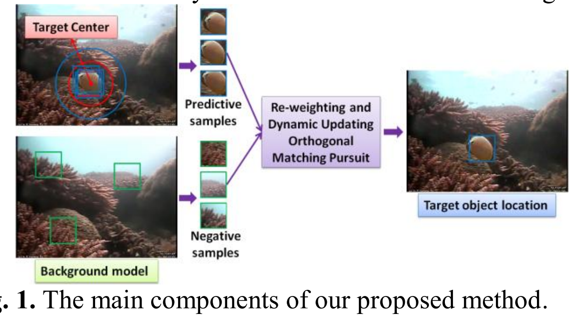
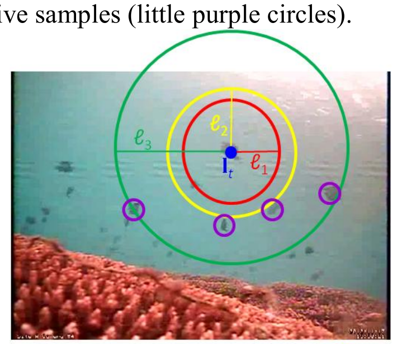


Fig. 1. The main components of our proposed method.

This paper is organized as follows. Section 2 briefly introduces the sample collection method using SuBM. Section 3 describes the sparse sample representation method using RwDuOMP. Section 4 presents the experimental results, and Section 5 provides the conclusions drawn.

2. SAMPLE COLLECTION METHOD BASED ON SPEED-UP BACKGROUND MODELING

The CT [5] method, with an appearance model based on features extracted from the compressed domain, is proposed and successfully applied in real-time object tracking. Three kinds of samples (positive, negative, and predictive) are collected. A set of positive samples $\boldsymbol{P} = \{\mathbf{Z} \mid \|\mathbf{l}(\mathbf{Z}) - \mathbf{l}_t\| < \ell_1\}$ is acquired near the current target object location, where $\mathbf{Z}$ is the sample, $\mathbf{l}_t$ is the tracking location at the t^{th} frame and ℓ_1 is the search radius for drawing positive samples. A set of negative samples $\boldsymbol{N} = \{\mathbf{Z} \mid \ell_2 < \|\mathbf{l}(\mathbf{Z}) - \mathbf{l}_t\| < \ell_3\}$ far from the center of the target object is obtained, where $\ell_1 < \ell_2 < \ell_3$. To predict the target object location, a set of predictive samples $\boldsymbol{Y} = \{\mathbf{Z} \mid \|\mathbf{l}(\mathbf{Z}) - \mathbf{l}_{t-1}\| < \ell_4\}$ is acquired around the current target object location. The predictive sample with the maximum classification score using the naïve Bayes classifier is assigned as the target object in the next frame.
However, the sample collection method in CT is likely to cause the target fish to gradually drift away in several situations, such as when the target fish move rapidly out of the assigned tracking window or when similar fish and background scenes are present in the next frame. In these situations, the two sets of predictive and negative samples are likely to include noisy samples. Figure 2 illustrates the noisy negative samples (little purple circles).



noisy negative samples : ○

Fig. 2. The noisy negative samples (little purple circles). ℓ_1 is the radius of red circle, ℓ_2 is the radius of yellow circle, and ℓ_3 is the radius of green circle.

2.1. Sample Collection Method

To overcome noisy sample collection issue when using CT, an enhanced sample collection method is developed. The procedures used to collect the predictive and negative samples in CT are replaced using SuBM. The negative samples are collected from the built background model (BM). The predictive samples are collected from the range (the radius is ℓ_4), where the center of the tracking window is located at the tracking result using SuBM. The collection of negative samples using our sample collection method efficiently avoids the acquisition of noisy samples, i.e., similar fish. Our method for obtaining predictive samples also overcomes the issues caused by fast-moving fish.

2.2. Speed-up Background Modeling Method Using Nonparametric Histogram

Each pixel of the GMM uses a mixture of K Gaussian distributions. The probability function of the pixel value at the t^{th} frame is

$$\mathcal{P}(x_t) = \sum_{j=1}^{K} w_{j,t} * \mathcal{F}\left(x_t, \mu_{j,t}, \sigma_{j,t}^2\right) \quad (1)$$

where $\mathcal{F}$ is a Gaussian probability density function with the mean $\mu_{j,t}$ and the variance $\sigma_{j,t}^2$, and $w_{j,t}$ is the weight of the j^{th} Gaussian distribution. However, computing the weight $w_{j,t}$ and the parameters $\mu_{j,t}$ and $\sigma_{j,t}^2$ cost a lot of time. In this study, a SuBM method using nonparametric histogram concept is adopted to improve the efficiency of the GMM. Each pixel of the SuBM at the t^{th} frame is represented by an H-bin histogram

$$\mathcal{P} = \left\{\mathcal{P}_{h,t} \mid h \epsilon \{0, \dots, H-1\}\right\} \quad (2)$$

where h is the bin index and $\mathcal{P}_{h,t}$ is the background probability of the h^{th} bin. We define Δb as the bin width, and the intensity range of the h^{th} bin is $[b_h, b_{h+1})$, where $b_h = b_0 + h \cdot \Delta b$. Then, a mapping function $B(x_t): X \to B$ is proposed to map the pixel values $X = \{0, \dots, 255\}$ to the bin indices $B = \{0, \dots, H-1\}$. The steps of the SuBM method are summarized as follows:

1. Input 1^{th} image frame.
2. Initialize background model. The first value of the nonparametric histogram background model $\mathcal{P}$ is set by
$$\mathcal{P}_{h,t=0} = \begin{cases} 1 & \text{if } h = B(x_{t=0}) \\ 0 & \text{otherwise} \end{cases}, \forall h \in \{0, \dots, H-1\} \quad (3)$$
3. Update background model. The pixel value at the t^{th} frame and the background model $\mathcal{P}$ are updated by first decreasing probability and then increasing probability.
   (a) Decrease probability: each bin value $\mathcal{P}_{h,t-1}$ at the $(t\text{-}1)^{th}$ frame is first decreased by
   $$\mathcal{P}_{h,t} = \delta \mathcal{P}_{h,t-1}, \forall h, \quad (4)$$
   where $\delta \in [0,1]$ is a decreasing parameter.
   (b) Increase probability: for the $B(x_t)^{th}$ bin value $\mathcal{P}_{B(x_t),t}$ at the t^{th} frame is then increased by
   $$\mathcal{P}_{B(x_t),t} = \mathcal{P}_{B(x_t),t} + \upsilon \quad (5)$$
   where $\upsilon \in [0,1]$ is an increasing parameter.
4. Extract foreground pixels. A binary pixel label is assigned by
$$L(x_t) = \begin{cases} 0 & \text{if } \mathcal{P}_{B(x_t),t} \geq \mathcal{T} \\ 1 & \text{otherwise} \end{cases} \quad (6)$$
where $\mathcal{T}$ is a threshold. $L(x_t) = 0$ means the pixel belongs to the background and $L(x_t) = 1$ denotes the pixel belongs to the foreground.

3. SPARSE SAMPLE REPRESENTATION METHOD USING RE-WEIGHTING AND DYNAMICALLY UPDATING ORTHOGONAL MATCHING PURSUIT

3.1. Sparse Sample Representation Method

In SRC [6], a testing image is used to represent a sparse linear combination of all training images, and the weighting coefficients are calculated via l_1-norm minimization. In this study, in order to obtain a sparse solution, a sparse sample representation method based on SRC is developed. Each predictive sample is used to represent a sparse linear combination of the positive and negative samples. We assign a sparse rate $\mathcal{R} = p/(p+n)$, where p is the number of positive samples and n is the number of negative samples. When $p << n$, the value of $\mathcal{R}$ is small, which also implies that $\mathcal{R}$ is sparse. We set $\boldsymbol{S} = [\boldsymbol{P}, \boldsymbol{N}] \in \mathbb{R}^{m\times(p+n)}$ as the concatenation of the positive and negative samples. $\boldsymbol{P} = [\boldsymbol{P}_1, \boldsymbol{P}_2, \dots, \boldsymbol{P}_p] \in \mathbb{R}^{m\times p}$ is the set of positive samples, where $\boldsymbol{P}_i$ is an m-dimensional vector stretched by the i^{th} positive sample. $\boldsymbol{N} = [\boldsymbol{N}_1, \boldsymbol{N}_2, \dots, \boldsymbol{N}_n] \in \mathbb{R}^{m\times n}$ is the set of negative samples, where $\boldsymbol{N}_i$ is an m-dimensional vector stretched by the i^{th} negative sample. A predictive sample $\boldsymbol{y} \in \mathbb{R}^m$ could be represented as a linear combination of $\boldsymbol{S}$. Figure 3 illustrates the representation of the predictive sample. The linear equation can be written as $\boldsymbol{y} = \boldsymbol{S\varphi}$, where $\boldsymbol{\varphi} = [\varphi_1; \varphi_2; \dots; \varphi_p; \varphi_{p+1}; \dots; \varphi_{p+n}] \in \mathbb{R}^{p+n}$ are weighting coefficients. The significant values of the weighting coefficients $\varphi_1; \varphi_2; \dots; \varphi_p$ can be calculated and values of the weighting coefficients $\varphi_{p+1}; \dots; \varphi_{p+n}$ are trivial, i.e., nearly equal to zero. Then, the weights of the predictive samples are computed using our proposed RwDuOMP method.

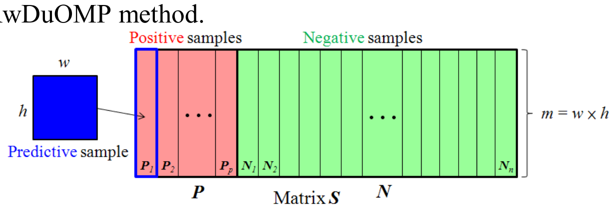


Fig. 3. The representation of the predictive sample.

3.2. Re-weighting and Dynamically Updating Orthogonal Matching Pursuit Method

OMP [8], a greedy algorithm based on vector projection, is proposed to reconstruct sparse signal. In order to improve the performance of sparse signal reconstruction, a RwDuOMP method based on OMP is proposed. This method includes three concepts: picking extra samples, re-weighting the picked samples, and dynamically updating negative samples. Firstly, supposing that the signal sparsity is s_o, we assign the extra samples $s_e = s_o + e$, where $e \geq 1$. Picking extra samples ensures that more positive samples (significant samples) can be picked. In addition, in order to raise the influence of positive samples and reduce the influence of negative samples, the concept of re-weighting picked samples is proposed. We re-weight the weighting coefficients so that the weighting coefficients of the picked positive samples are increased, and the weighting coefficients of the picked negative samples are decreased at the t^{th} frame. The original weight of each predictive sample is $w_i(\boldsymbol{y}) = \sum_{i=1}^{a} \alpha_i + \sum_{j=1}^{b} \beta_j$, where a is the number of picked positive samples, and α_i is the i^{th} weighting coefficient of the picked positive samples. b is the number of picked negative samples, and β_j is the j^{th} weighting coefficient of the picked negative samples. Then, the weight of each predictive sample using our method is $w_i(\boldsymbol{y}) = \gamma \sum_{i=1}^{a} \alpha_i + \xi \sum_{j=1}^{b} \beta_j$, where γ is the increasing parameter ($\gamma > 1$), and ξ is the decreasing parameter ($\xi < 1$). Re-weighting picked samples ensures that positive samples have a higher influence and negative samples have a lower influence. Moreover, in order to further eliminate the noise of negative samples, dynamically updating of negative samples is proposed. The picked negative samples at the $(t-1)^{th}$ frame denotes they are similar to predictive samples. At the t^{th} frame, these picked negative samples are eliminated and compensated for by collecting new negative samples. Dynamically updating negative samples can efficiently eliminate the noise of negative samples.

The steps of the SSCR method are summarized as follows.

1. Input t^{th} image frame.
2. Implement SuBM method to construct the BM, and determine the foreground target object to obtain the target object center $\mathbf{l}_{tc}$.
3. Collect a set of positive samples $\boldsymbol{P} = \{\mathbf{Z} | \|\mathbf{l}(\mathbf{Z}) - \mathbf{l}_t\| < \ell_1\}$ and a set of negative samples $\boldsymbol{N} = \{\mathbf{Z} | \mathbf{Z} \in \text{BM}\}$.
4. Extract the Haar-like features from the positive and negative sample sets.
5. Collect a set of predictive samples $\boldsymbol{Y} = \{\mathbf{Z} | \|\mathbf{l}_{t+1}(\mathbf{Z}) - \mathbf{l}_{tc}\| < \ell_4\}$.
6. Set $\boldsymbol{S} = [\boldsymbol{P}, \boldsymbol{N}] \in \mathbb{R}^{m\times(p+n)}$ as a matrix of the positive and negative samples, and a predictive sample $\boldsymbol{y} \in \mathbb{R}^m$, where $\boldsymbol{y} = \boldsymbol{S\varphi}$.
7. Compute the weight $w_i(\boldsymbol{y}) = \gamma \sum_{i=1}^{a} \alpha_i + \xi \sum_{j=1}^{b} \beta_j$ of each predictive sample $\boldsymbol{y}$ using the RwDuOMP method.
8. Label $\boldsymbol{y}$ by identity $(\boldsymbol{y}) = \arg\{\max_i w_i(\boldsymbol{y})\}$, and output tracking location $\mathbf{l}_{t+1}$.

4. EXPERIMENTAL RESULTS

Several challenging real-world underwater sequences are used to evaluate the performance of our proposed tracking method (SSCR). These video sequences are sampled to 640×480 pixels with 24-bit RGB bitmaps and a frame rate of 24 fps. The total number of evaluated frames is 8,952. The initial window and other parameters ($\ell_1 = 5$, $\ell_4 = 10$, $p = 5$, $n = 45$, $e = 5$, $\gamma = 1.1$, $\xi = 0.9$) are assigned the same values in CT and SSCR. A metric is the tracking success rate (TSR)(%), which is used to evaluate the proposed method with existing algorithms. The TSR score is determined by $\frac{\text{area}(\text{ROI}_T \cap \text{ROI}_G)}{\text{area}(\text{ROI}_T \cup \text{ROI}_G)}$, where ROI_T is the tracking target bounding box and ROI_G is the ground truth bounding box. The tracking result is considered a success if the score is larger than 0.5 in one frame. In addition, our proposed ReDuOMP method is also evaluated with OMP, ROMP, and CoSaMP. Table 1 shows the TSR of all methods. Figure 4 shows the tracking result of all methods using several different scenarios from real-world underwater sequences. The purple box is the SSCR_RwDuOMP result, the yellow box is the SSCR_OMP result, the red box is the SSCR_ROMP result, the green box is the SSCR_CoSaMP result, and the blue box is the CT result. In most video sequences, SSCR obtains better tracking results than CT, particularly in challenging situations, such as fast-moving fish and the presence of similar fish.

Table 1. Tracking Success Rate (TSR)(%): (1) SSCR_RwDuOMP; (2) SSCR_OMP; (3) SSCR_ROMP; (4) SSCR_CoSaMP; (5) CT.

	(1)	(2)	(3)	(4)	(5)
TSR	83.6%	79.2%	73.6%	77.6%	68.7%

Fig. 4. Tracking results: purple box (SSCR_RwDuOMP), yellow box (SSCR_OMP), red box (SSCR_ROMP), green box (SSCR_CoSaMP), and blue box (CT).

We also compare the performance between the SuBM and the GMM. The parameters (Δb=256$/H$, $\mathcal{T}$=0.25) are adopted in the SuBM. In the experimental comparisons, the average speed of the SuBM is 2.61 times faster than the GMM when H=8, while the average speed of the SuBM is 2.46 times faster than the GMM when H=16.

5. CONCLUSION

In this study, a SSCR based on CT and SRC was developed for real-world fish tracking. It included two procedures: sample collection and sparse sample representation. In the sample collection procedure, three sets of positive, negative, and predictive samples were collected. A SuBM method using nonparametric histogram concept was proposed to collect more satisfactory predictive and negative samples and efficiently accelerate the tracking speed in the sample collection procedure. In the sparse sample representation procedure, each predictive sample was represented as a sparse linear combination of all the positive and negative samples. The weights of predictive samples were computed using our proposed RwDuOMP method, and the predictive sample with the maximum weight was regarded as the target object tracking result. The RwDuOMP, including three concepts (picking extra samples, re-weighting picked samples, and dynamically updating negative samples), was able to improve the performance of the sparse signal reconstruction. Numerous experiments that considered challenging sequences from different real-world underwater scenes demonstrated that the SSCR achieved a better accuracy than CT. Our RwDuOMP method also obtains better tracking results than OMP, ROMP, and CoSaMP, and our SuBM method was faster than the GMM.

6. ACKNOWLEDGMENTS

The research is funded by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. MOST104-2221-E-007-096-MY2. We thank the Third Nuclear Power Plant of Taiwan Power Company and National Museum of Marine Biology & Aquarium, Taiwan for logistical support.

7. REFERENCES

[1] Y. Tian, R. Feris, H. Liu, A. Humpapur, and M.T. Sun, "Robust Detection of Abandoned and Removed Objects in Complex Surveillance Videos," *IEEE Trans. On System, Man, and Cybermetics, Part C: Applications and Reviews*, pp. 1-12, vol. 99, 2011.

[2] S. Brutzer, B. Hoferlin, and G. Heidemann, "Evaluation of Background Subtraction Techniques for Video Surveillance," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1937-1944, 2011.

[3] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 246-252, vol. 2, 1999.

[4] E.J. Candes, M.B. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Processing Magazine*, pp. 21-30, 2008.

[5] K. Zhang, L. Zhang, and M.H. Yang, "Real-time Compressive Tracking," *Proceedings of European Conference on Computer Vision*, pp. 864-877, 2012.

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

[7] E. Strandell, S. Tilak, H.M. Chou, Y.T. Wang, F.P. Lin, P. Arzberger, T. Fountain, T.Y. Fan, R.Q. Jan, and K.T. Shao, "Data Management at Kenting's Underwater Ecological Observatory," *International Conference on Intelligent Sensors, Sensor Networks, and Information*, pp. 715-720, 2007.

[8] J.A. Tropp, and A.C. Gilbert, "Signal Recovery From Measurements via Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory*, pp. 4655-4666, vol. 53, no. 12, 2007.

[9] D. Needell, and J.A. Tropp, "Signal Recovery From Incomplete and Inaccurate Measurements via Regularized Orthogonal Matching Pursuit," *IEEE Journal of Selected Topics in Signal Processing*, pp. 310-316, vol. 4, no. 2, 2010.

[10] D. Needell, and J.A. Tropp, "CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples," *Applied and Computational Harmonic Analysis*, pp. 301-321, vol. 26, no. 3, 2009.