

Received 6 October 2022, accepted 18 November 2022, date of publication 21 November 2022,
date of current version 2 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3224027

RESEARCH ARTICLE

Maximum Margin Correlation Filter Based on Data Spatial Distribution Information

QI JIANG¹, JIA LUO¹, GANG ZHOU¹, AND SAMYA BAGCHI^{2,3}

¹School of Computer Science, Sichuan University Jinjiang College, Meishan 620860, China

²Computational Learning Systems Laboratory, School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes, SA 5095, Australia

³AniXR Corporation, Mawson Lakes, SA 5095, Australia

Corresponding author: Jia Luo (1529900947@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51208034, in part by the Science and Technology Planning Project of Sichuan Provincial Department of Science and Technology under Grant 2021YJ0315, in part by the Science and Technology Planning Project of Meishan Science and Technology Bureau under Grant 2021YS05, and in part by the Open Fund Project of Key Laboratory of Pattern Recognition and Intelligent Information Processing in Sichuan Province Universities under Grant MSSB-2022.

ABSTRACT In the past decade, object localization and object classification using correlation filters, especially large margin correlation filters which combine with support vector machine (SVM), have become a hotspot. However, the large margin correlation filters do not consider the class distribution and the structural features within the class during training, which is easily affected by noise. This paper presents two methods to overcome this drawback: minimum class variance large margin correlation filter (MCVLMCF) and minimum class locality preserving variance correlation filter (MCLPVCF). First, the overall structure information of the target is obtained by the within-class scatter with MCVLMCF, and the spatial features of the sample are extracted by the intrinsic manifold structure of data with MCLPVCF. Then, we embed these two types of information into the optimal function of the large margin correlation filter, fuse the sample spatial features with the large margin principle and correlation filtering, and convert it to solve the filter in the frequency domain. Finally, object localization experiments in actual environments and classification experiments on different datasets demonstrate that our proposed methods can adapt to complex object changes and achieve better performances than some state-of-the-art methods.

INDEX TERMS Correlation filter, support vector machine, within-class scatter, intrinsic manifold structure.

I. INTRODUCTION

Research on computer vision algorithms has been the focus of much activity in computer science. It covers a variety of application areas and academic disciplines, including target classification, target detection, digital image processing, geometric modeling, physics, and mathematics. Among these, target classification and target detection are significant computer vision research and application areas. They are frequently utilized in our lives for vehicle detection [1], pedestrian detection [2], object recognition [3], and face recognition [4].

In the past years, correlation filters have been investigated for target classification, detection, and tracking [5] because of

The associate editor coordinating the review of this manuscript and approving it for publication was Wenjie Feng.

their noise immunity, shift-invariance, and fast training. Correlation filters achieved good performance in many pattern recognition applications including face localization, pedestrian localization, and object tracking [5]. Many computer vision methods have been combined with correlation filters. Vander et al. [6] propose matched filter (MF) which is an earlier correlation filter algorithm. The construction method of the filter is the conjugate transpose after the two-dimensional Fourier transform of the training image. The structure of the MF is too simple, only the correlation output of training images is relatively accurate, and it is not practical enough, but the application of the correlation filter in image processing also provides new ideas for subsequent researchers. After a long period of improvement, the current correlation filter can be divided into two kinds: the synthetic

discriminant function filter and the optimal correlation output filter. When building the filter, the synthetic discriminant function constrains the output responses of the positive and negative samples to 1 and 0, respectively. The discriminant function is then used to assess whether the sample is positive or negative. Casacent et al. [7] propose the original synthetic discriminant function filter (SDF) which is the earliest proposed synthetic discriminant function filter, it sets a part of the positive data in the training set as high output values, sets other samples to low output values. However, it is essentially the weighted sum of multiple MFs, and the actual effect does not significantly exceed MF. Kumar et al. [8] propose the minimum variance synthetic discriminant function filter (MVSDF) which improved on SDF. MVSDF minimizes the correlation output variance when constructing the filter and adds constraints on noise output, suppresses noise and emphasizes low-frequency information, but it also suppresses peak output so that the correlation output is insufficiently noticeable. Mahalanobis et al. [9] propose the minimum average correlation energy (MACE) and the unconstrained minimum average correlation energy (UMACE) in literature [9], which increase the constraint on the correlation output, enhance the correlation output, but reinforce the correlation output to noise, and cause the anti-noise ability is insufficient, the output peak is drowned out in the noise output. Mahalanobis et al. [9] propose the maximum average correlation height (MACH) which simultaneously minimizes the mean square error of the noise and targets to construct a peak output with a high correlation height. Mahalanobis et al. [10] propose the unconstrained optimal trade-off synthetic discriminant function (UOTSDF) by fusing MVSDF and MACE and using a balancing parameter to combine the two filters. Kumar et al. [11], [12] propose minimum squared error synthetic discriminant functions (MSESDF) and Optimal Trade-off Synthetic Discriminant Functions (OTSDF) by utilizing MVSDF and MACE. Their correlation output is emphasized and the noise is suppressed, but performance is also significantly affected by parameter selection.

Another kind of correlation filter is the optimal correlation filter. This kind of correlation filter constructs a strong peak at the target or center position of the training images, and then learns a mapping from the training images to the desired output to calculate the filter. The average synthetic exact filter (ASEF), proposed by Bolme et al. [13], uses the construction of the target response for each training image and then averages it to obtain the final filter, and it obtains a weak filter for each image first and then merges it to find the strong filter. This algorithm performs well in the experiments of human eye localization and face recognition, because the deformation of these kinds of image data is generally small. However, the averaging approach loses a large amount of useful image information, and the performance is unsatisfactory when the image is more variable and the number of samples is small. Bolme et al. [14] propose the minimum output sum of squared error (MOSSE), which is constructed by solving

the sum of squares that minimizes the true correlation output and the desired correlation output of the training samples. MOSSE uses fewer training samples for training than ASEF, and it requires fewer demands on the training set, which speeds up the training process compared to ASEF. Most of the correlation filters are improved based on this algorithm.

Subsequently, Rodriguez et al. [15] changed the idea of improving the correlation filter, combined the correlation filter with support vector machine (SVM) [16], and propose the maximum margin correlation filter (MMCF). SVM is a classical supervised learning classifier that is widely used in pattern recognition tasks such as face recognition [17], and pedestrian detection [18]. MMCF maximizes the classification margin and minimizes the mean square error while utilizing SVM's generalization capabilities and correlation filters' classification capabilities. Thus, the obtained filter performs better than the previous traditional correlation filter algorithm in target detection and target recognition experiments. To improve MMCF, Boddeti et al. [19] propose the maximum margin vector correlation filter (MMVCF), and Fernandez et al. [20] propose the zero-aliasing maximum margin correlation filter (ZAMMCF). Compared with MMCF, MMVCF uses vector features, and ZAMMCF introduces a new zero aliasing constraint to eliminate the aliasing problem of correlation filters.

Although compared to other correlation filters, MMCF and its subsequently improved algorithms have achieved excellent performance, these methods still have some shortcomings. First, these algorithms do not consider the overall distribution information of the samples when constructing the filters, and only apply the sample points on the boundary; Second, they ignore the fact that the structural information between samples is also an important factor in filter construction. These drawbacks make the filter susceptible to noise during training, which can easily be incorrectly classified when the non-targets are more similar to the targets. To address the problems, we propose two new methods. For the first problem, we propose the minimum class variance large margin correlation filter (MCVLMCF) using fisher linear discrimination [21], [22] and SVM. For the second problem, by using locality preserving projections [23], [24], [25] (LPP) and SVM, we propose the minimum class locality preserving variance correlation filter (MCLPVCF). The main contributions of our work can be summarized as follows: (1) We introduce the within-class scatter into the large margin correlation filter, so that the algorithm can take into account the boundary samples and the overall distribution structure of the samples during training. (2) Then, using the sample manifold structure, the local preserving scatter is combined with a large margin correlation filter, and the internal structure information of the sample is fully considered. (3) Finally, we investigate the influence of parameters on the proposed algorithms through some experiments, and compare the proposed algorithms with other algorithms through object localization and object classification experiments.

II. MAXIMUM MARGIN CORRELATION FILTER

Before the proposal of MMCF, the improvement method of the traditional correlation filter was based on the optimization of the correlation output, and did not consider combining other classifiers to improve its discrimination ability. Rodriguez et al. [15] extend the correlation filter and fuse SVM with the correlation filter. MMCF combines the maximum margin principle of SVM with correlation filter, and uses the excellent classification and generalization capabilities of SVM to further improve the performance of the filter.

A. CORRELATION FILTER

Many correlation filter designs can be interpreted as optimizing a distance metric between an ideal desired correlation output for an input image and the correlation output of the training images with the filter template [26]. Given a dataset $U = \{\mathbf{x}_i | i = 1, \dots, N\} \subset \mathbb{R}^{d \times N}$, the definition of the correlation filter is as follows:

$$\mathbf{w} = \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{i=1}^N \|\mathbf{w} \otimes \mathbf{x}_i - \mathbf{g}_i\|_2^2 \quad (1)$$

In Eq.1, \mathbf{w} is the filter vector, \mathbf{x}_i is the input image vector, and \mathbf{g}_i is the ideal desired output vector. The definitions of \mathbf{g}_i for different filters are not exactly the same, but basically the center position has a high value, and the rest of the positions have low values. \otimes represents the cross-correlation operation of the image.

B. SVM

In image processing, SVM requires feature vectors of the same size for training and prediction, which can lexicographically scan image pixel values to form feature vectors, and then label the feature vectors in the training image [27]. Given a dataset $U = \{\mathbf{x}_i | i = 1, \dots, N\} \subset \mathbb{R}^{d \times N}$, and the sample label $y_i \in \{1, -1\}$, the SVM maximizes the minimum L-2 norm distance (also known as the margin) between the hyperplane and any data sample by solving this hyperplane:

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \quad i = 1, \dots, N \end{aligned} \quad (2)$$

In Eq.2, \mathbf{w} is the normal vector of the classification hyperplane, T represents transposition, C is a given constant that defines the cost of the errors after the classification, ξ_i is the slack variable, and b is the bias or offset from the origin.

C. CORRELATION FILTER THAT INTRODUCE SUPPORT VECTORS

By combining SVM with correlation filter, MMCF not only minimizes the mean squared error, but also simultaneously maximizes the minimum Euclidean distance (margin) between the hyperplane and the data points. The function of

MMCF is as follows:

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \left(\mathbf{w}^T \mathbf{w} + C \sum_{i=1}^T \xi_i \right) + (1 - \lambda) \sum_{i=1}^N \|\mathbf{w} \otimes \mathbf{x}_i - \mathbf{g}_i\|_2^2 \\ \text{s.t.} \quad & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq c_i - \xi_i, \xi_i \geq 0, \quad i = 1, \dots, N \end{aligned} \quad (3)$$

In Eq.3, $\lambda \in (0, 1)$ is balance parameter, $c_i = 1$ for true-class images and $c_i = 0$ for false-class images. MMCF takes $\mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i$ as the classification criterion and $\sum_{i=1}^N \|\mathbf{w} \otimes \mathbf{x}_i - \mathbf{g}_i\|_2^2$ as the localization criterion. Compared with traditional correlation filters, MMCF has significantly improved classification and localization capabilities.

III. CORRELATION FILTER BASED ON MINIMUM WITHIN-CLASS VARIANCE

MMCF, MMVCF, and ZAMMCF introduce the principle of maximum margin into the correlation filter. Compared with traditional correlation filters, these filters perform better in object detection and object recognition experiments. However, the SVM part of MMCF only considers the sample points on the boundary when establishing the classification hyperplane, ignoring the distribution information of various sample data. Thus, the classification hyperplane obtained by SVM is not consistent with the actual situation of the class distribution. Consequently, MMCF applies only boundary sample points during training.

A. MINIMUM CLASS VARIANCE SUPPORT VECTOR MACHINE

In the classification problem, SVM attempts to obtain the classification hyperplane with the maximum margin, where the margin is defined as the minimum distance from the class boundary sample to the hyperplane, which causes the aforementioned problem. Stefanos et al. [28] combined Fisher's linear discriminant analysis (FLDA) [29] and SVM, and proposed the minimum within-class variance support vector machine (MCSVSM). Suppose a training sample set $U = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \subset \mathbb{R}^{d \times N}$, which belong to two categories $\{A_+, A_-\}$, MCSVSM's optimization problem is defined as:

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \mathbf{w}^T \mathbf{S}_\omega \mathbf{w} + C \sum_{i=1}^N \xi_i \\ \text{s.t.} \quad & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, \quad i = 1, \dots, N \end{aligned} \quad (4)$$

In Eq.4, \mathbf{S}_ω is the within-class scatter matrix, which is a symmetric positive semi-definite matrix, and is defined as follows:

$$\begin{aligned} \mathbf{S}_\omega = \quad & \sum_{\mathbf{x} \in A_-} (\mathbf{x} - \mathbf{m}_{A_-}) (\mathbf{x} - \mathbf{m}_{A_-})^T \\ & + \sum_{\mathbf{x} \in A_+} (\mathbf{x} - \mathbf{m}_{A_+}) (\mathbf{x} - \mathbf{m}_{A_+})^T \end{aligned} \quad (5)$$

In Eq.5, \mathbf{m}_{A_-} and \mathbf{m}_{A_+} are mean sample vectors for the classes A_+ and A_- , respectively. If the input image dimension

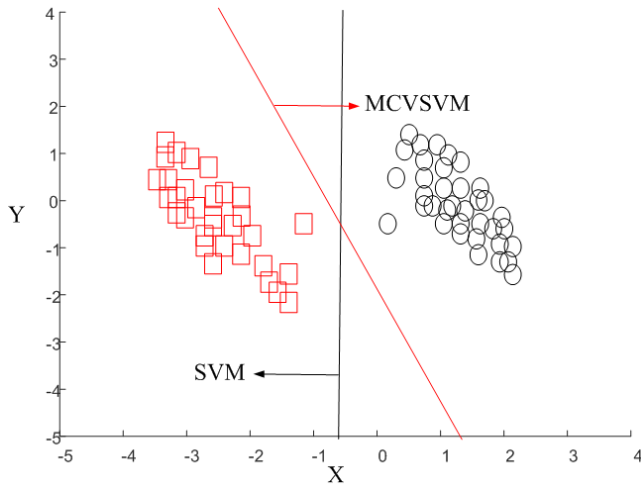


FIGURE 1. Classification hyperplane for SVM and MCVSVM.

is d , the size of S_ω is $d \times d$. The within-class scatter matrix is used to describe the distribution within the sample data class. The distance from the sample to the hyperplane in the SVM is the Euclidean distance. In contrast, after the within-class scatter is introduced, MCVSVM applies the Mahalanobis distance [29].

Fig.1 shows the classification hyperplane generated by SVM and MCVSVM, in which circles and rectangles represent two classes of artificial data. SVM only takes into consideration the Euclidean distance between the two classes of samples and is affected by the two sample points in the middle of the dataset. The obtained decision hyperplane does not conform to the spatial distribution of the samples. However, MCVSVM introduces the within-class scatter matrix to incorporate the overall spatial distribution information for training, and it does not be disturbed by individual special sample points.

It can be found that the classification hyperplane of MCVSVM is more in line with the overall distribution of the data, and better solves the problem that SVM does not consider sample distribution information. Therefore, MCVSVM is more robust than SVM and has a better generalization performance. Our improvements to the large interval correlation filter are inspired by this approach.

B. MINIMUM CLASS VARIANCE LARGE MARGIN CORRELATION FILTER

To address the above-mentioned problem of the correlation filter, we propose MCVLMCF inspired by [28]. Referring to Eq.3, the MCVLMCF model can be established as follows:

$$\begin{aligned} \min_{w,b} & \lambda \left(w^T S_\omega w + C \sum_{i=1}^N \xi_i \right) + (1 - \lambda) \sum_{i=1}^N \|w \otimes x_i - g_i\|_2^2 \\ \text{s.t.} & y_i (w^T x_i + b) \geq c_i - \xi_i, \xi_i \geq 0, i = 1, \dots, N \end{aligned} \quad (6)$$

In Eq.6, λ is used to balance the proportion of the classification criterion part and the localization criterion part in

MCVLMCF, and the value range is (0,1]. For example, when $\lambda = 1$, it is equivalent to ignoring the positioning standard part, and only MCVSVM is applied. The specific structural form of S_ω is shown in Eq.5, which uses the scatter of the data within the class to obtain the distribution information of the data samples.

Inspired by MCVSVM, MCVLMCF introduces the sample scatter within the class, takes full account of the distribution information of the data, and makes the training of MCVLMCF more sufficient, which is not easily affected by the noise. In addition, the performance of the filter can be improved by combining the advantages of the correlation filter and the maximum margin principle of SVM.

C. THE SOLUTION OF MCVLMCF

As shown in Eq.6 that MCVLMCF is a combination of two optimization problems, it is difficult and time-consuming to solve MCVLMCF directly. According to the correlation theorem [30], the cross-correlation operation in the correlation filter can be transformed into multiplication in the frequency domain through the Fourier transform, which can improve the operation speed. Therefore, referring to the solution method in the literature [15], we convert the model to the frequency domain and solve it using the Wolf dual problem.

To facilitate the calculation, we convert the localization and classification criterion of MCVLMCF respectively. The localization criterion is transformed as follows:

$$\begin{aligned} \sum_{i=1}^N \|w \otimes x_i - g_i\|_2^2 &= \frac{1}{d} \sum_{i=1}^N \left\| \hat{X}_i^* \hat{w} - \hat{g}_i \right\|_2^2 \\ &\propto \sum_{i=1}^N \left\| \hat{X}_i^* \hat{w} - \hat{g}_i \right\|_2^2 \end{aligned} \quad (7)$$

In Eq.7, \hat{X}_i is a diagonal matrix whose elements on the diagonal are all the elements of \hat{x}_i . $\hat{x}_i, \hat{w}, \hat{g}_i$ are the input image vector x_i , filter vector w , and desired output vector g_i which are converted to the frequency domain by the Fourier transform, respectively. We ignore $1/d$ because it is a constant and has no effect on the filter.

\hat{g}_i in the frequency domain can be rewritten as:

$$\hat{g}_i = \mathbf{1} x_i^T w = \frac{1}{d} \mathbf{1} \hat{x}_i^\dagger \hat{w} \quad (8)$$

In Eq.8, \dagger represents the conjugate transpose, $\mathbf{1} = [1 \dots 1]^T$, which is a d -dimensional vector. Substitute Eq.8 into Eq.7, it can be changed as follows:

$$\begin{aligned} \sum_{i=1}^N \left\| \hat{X}_i^* \hat{w} - \hat{g}_i \right\|_2^2 &= \sum_{i=1}^N \left(\hat{w}^\dagger \hat{X}_i \hat{X}_i^* \hat{w} - 2 \hat{w}^\dagger \hat{X}_i \hat{g}_i + \hat{g}_i^\dagger \hat{g}_i \right) \\ &= \sum_{i=1}^N \left(\hat{w}^\dagger \hat{X}_i \hat{X}_i^* \hat{w} - \frac{2}{d} \hat{w}^\dagger \hat{X}_i \mathbf{1} \hat{x}_i^\dagger \hat{w} + \frac{1}{d^2} \hat{w}^\dagger \hat{x}_i \mathbf{1}^\dagger \mathbf{1} \hat{x}_i^\dagger \hat{w} \right) \\ &= \sum_{i=1}^N \left(\hat{w}^\dagger \hat{X}_i \hat{X}_i^* \hat{w} - \frac{2}{d} \hat{w}^\dagger \hat{x}_i \hat{x}_i^\dagger \hat{w} + \frac{1}{d} \hat{w}^\dagger \hat{x}_i \hat{x}_i^\dagger \hat{w} \right) \end{aligned}$$

$$\begin{aligned}
&= \widehat{\mathbf{w}}^\dagger \left(\sum_{i=1}^N \widehat{\mathbf{X}}_i \widehat{\mathbf{X}}_i^* - \frac{1}{d} \sum_{i=1}^N \widehat{\mathbf{x}}_i \widehat{\mathbf{x}}_i^\dagger \right) \widehat{\mathbf{w}} \\
&= \widehat{\mathbf{w}}^\dagger \widehat{\mathbf{Z}} \widehat{\mathbf{w}} \tag{9}
\end{aligned}$$

In Eq.9, $\widehat{\mathbf{Z}} = \sum_{i=1}^N \widehat{\mathbf{X}}_i \widehat{\mathbf{X}}_i^* - 1/d \sum_{i=1}^N \widehat{\mathbf{x}}_i \widehat{\mathbf{x}}_i^\dagger$. In addition, because the dimension d of the image $\widehat{\mathbf{x}}_i$ is generally very large, borrowing from the processing method in the literature [15], to reduce the computational complexity, $1/d \sum_{i=1}^N \widehat{\mathbf{x}}_i \widehat{\mathbf{x}}_i^\dagger$ is ignored, which does not affect the final result.

We can formulate the MCVSVM in the frequency domain using the fact that inner products are only scaled by $1/d$ [31]. Thus, Eq.4 can be rewritten as follows:

$$\begin{aligned}
\min_{\mathbf{w}, b, \xi} \quad & \widehat{\mathbf{w}}^\dagger \widehat{\mathbf{S}}_\omega \widehat{\mathbf{w}} + C \sum_{i=1}^N \xi_i \\
s.t. \quad & y_i (\widehat{\mathbf{w}}^\dagger \widehat{\mathbf{x}}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, N \tag{10}
\end{aligned}$$

In Eq.10, $b' = b \times d$, $\widehat{\mathbf{S}}_\omega = \sum_{\widehat{\mathbf{x}} \in A_-} (\widehat{\mathbf{x}} - \widehat{\mathbf{m}}_{A_-}) (\widehat{\mathbf{x}} - \widehat{\mathbf{m}}_{A_-})^T + \sum_{\widehat{\mathbf{x}} \in A_+} (\widehat{\mathbf{x}} - \widehat{\mathbf{m}}_{A_+}) (\widehat{\mathbf{x}} - \widehat{\mathbf{m}}_{A_+})^T$, and the definition of this matrix is similar to the previous definition.

We can now express the MMCF multi-criteria shown in Eq. 6 in the frequency domain as follows:

$$\begin{aligned}
\min_{\mathbf{w}, b'} \quad & \lambda \left(\widehat{\mathbf{w}}^\dagger \widehat{\mathbf{S}}_\omega \widehat{\mathbf{w}} + C \sum_{i=1}^N \xi_i \right) + (1 - \lambda) \widehat{\mathbf{w}}^\dagger \widehat{\mathbf{Z}} \widehat{\mathbf{w}} \\
s.t. \quad & y_i (\widehat{\mathbf{w}}^\dagger \widehat{\mathbf{x}}_i + b') \geq c_i - \xi_i, \xi_i \geq 0, i = 1, \dots, N \tag{11}
\end{aligned}$$

Eq.11 can be solved by the Lagrange dual problem $L(\widehat{\mathbf{w}}, b', \xi, \alpha_i, \beta_i)$:

$$\begin{aligned}
L = \lambda \left(\widehat{\mathbf{w}}^\dagger \widehat{\mathbf{S}}_\omega \widehat{\mathbf{w}} + C \sum_{i=1}^N \xi_i \right) + (1 - \lambda) \widehat{\mathbf{w}}^\dagger \widehat{\mathbf{Z}} \widehat{\mathbf{w}} \\
- \sum_{i=1}^N \alpha_i \left[y_i (\widehat{\mathbf{w}}^\dagger \widehat{\mathbf{x}}_i + b') - c_i + \xi_i \right] - \sum_{i=1}^N \beta_i \xi_i \tag{12}
\end{aligned}$$

In Eq.12, α_i and β_i are the Lagrange multipliers. The Karush–Kuhn–Tucker (KKT) conditions imply that, for Eq.12, the following holds:

$$\begin{aligned}
\frac{\partial L}{\partial \widehat{\mathbf{w}}} = 0 &\Rightarrow \widehat{\mathbf{w}} = \frac{1}{2} [\lambda \widehat{\mathbf{S}}_\omega + (1 - \lambda) \widehat{\mathbf{Z}}]^{-1} \sum_{i=1}^N \alpha_i y_i \widehat{\mathbf{x}}_i \\
\frac{\partial L}{\partial b'} = 0 &\Rightarrow \sum_{i=1}^N \alpha_i y_i = 0 \\
\frac{\partial L}{\partial \xi_i} = 0 &\Rightarrow \sum_{i=1}^N \alpha_i = \lambda C - \sum_{i=1}^N \beta_i \tag{13}
\end{aligned}$$

By substituting the results in Eq.13 into Eq.12, the dual problem of MCVLMCF can be obtained:

$$\begin{aligned}
\min_{\alpha} \quad & \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \widehat{\mathbf{x}}_i^\dagger [\lambda \widehat{\mathbf{S}}_\omega + (1 - \lambda) \widehat{\mathbf{Z}}]^{-1} \widehat{\mathbf{x}}_j \\
& - \sum_{i=1}^N c_i \alpha_i \\
s.t. \quad & \sum_{i=1}^N \alpha_i y_i = 0; 0 \leq \alpha_i \leq \lambda C \tag{14}
\end{aligned}$$

In Eq.14, $\alpha = [\alpha_1, \dots, \alpha_N]^T$. Eq.14 can be solved by the sequential minimal optimization (SMO) [32]. SMO solves all Lagrange multipliers by heuristically searching for a pair of α_i , instead of solving the entire α vector at once. SMO breaks this large quadratic programming (QP) optimization problem into a series of the smallest possible QP problems so that it can be solved faster. The Lagrange multiplier vector α^* of the dual problem is solved, and the filter $\widehat{\mathbf{w}}^*$ of MCVLMCF in the frequency domain can be obtained using the following formula:

$$\widehat{\mathbf{w}}^* = [\lambda \widehat{\mathbf{S}}_\omega + (1 - \lambda) \widehat{\mathbf{Z}}]^{-1} \widehat{\mathbf{X}} \mathbf{Y} \alpha^* \tag{15}$$

In Eq.15, \mathbf{Y} is a diagonal matrix with y_i along the diagonal.

IV. MINIMUM CLASS LOCALLY PRESERVING VARIANCE GUIDED CORRELATION FILTER

When building the classification hyperplane, the SVM part of large margin correlation filters such as MMCF simply takes into account the margin between border samples and ignores the inherent manifold structure of the data space. Aiming at the drawback of this problem, we propose the minimum class locality preserving variance correlation filter (MCLPVCF). MCLPVCF, in contrast to other large margin correlation filters, fuses weighted adjacency graph of samples and introduces local preserving within-class scatter. It gives full consideration to structural information and manifold structure of samples, maximizes classification margin and optimizes correlation output, and then takes into consideration both class and structure information of samples during training.

A. MINIMUM CLASS LOCALITY PRESERVING VARIANCE SUPPORT VECTOR MACHINE

Aiming at the problem of SVM mentioned above, Wang et al. [33] use the locality preserving projections (LPP) [23], [24], [25], and the minimum class locality preserving variance support vector machine (MCLPV_SVM) is proposed in [33]. MCLPV_SVM optimization problem is defined as:

$$\begin{aligned}
\min_{\mathbf{w}, b} \quad & \mathbf{w}^T \mathbf{Z}_w \mathbf{w} + C \sum_{i=1}^N \xi_i \\
s.t. \quad & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, N \tag{16}
\end{aligned}$$

In Eq.16, \mathbf{Z}_w is the locally preserving within-class scatter matrix, which is used to describe the structural information

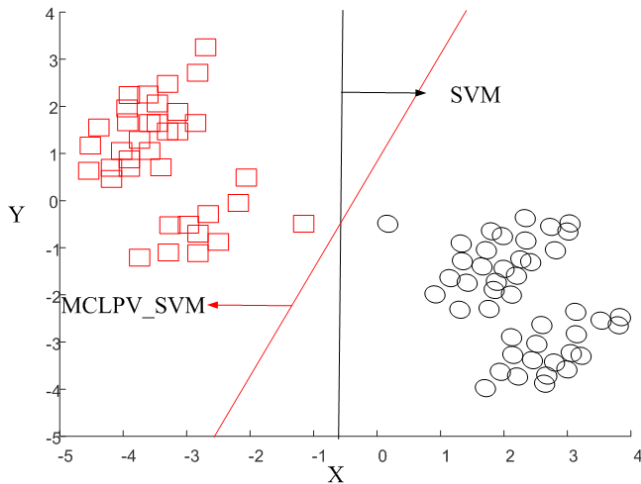


FIGURE 2. Classification hyperplane for SVM and MCLPV_SVM.

between data using the method of LPP. LPP is a data dimension reduction algorithm based on projection, which builds an adjacency graph containing structural information between the data. The locally preserving within-class scatter matrix is introduced by MCLPV SVM using this adjacency graph, with $K = 1, 2$ two-class samples, and its specific construction is as follows:

$$\mathbf{Z}_w = \sum_{K=1}^2 \mathbf{Z}_K \quad (17)$$

In Eq.17, \mathbf{Z}_K is the locally preserving within-class scatter matrix of the K -th class data. It is constructed as follows:

$$\mathbf{Z}_K = \mathbf{X}_K (\mathbf{D}^K - \mathbf{W}^K) \mathbf{X}_K^T \quad (18)$$

In Eq.18, \mathbf{X}_K is the input image matrix of the K -th class data, \mathbf{D}^K is a diagonal matrix whose elements D_{ij}^K is constructed as $D_{ij}^K = \sum_j W_{ij}^K$, W_{ij}^K is the element of the \mathbf{W}^K , \mathbf{W}^K is the adjacency graph weight matrix. \mathbf{x}_{Ki} and \mathbf{x}_{Kj} are the data points of the K -th class data, G is the adjacency graph of the dataset D , which denote the local manifold structure, and $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix of G . G has a total of N nodes (that is, data points), if node i is in the k nearest neighbors of node j , then $\mathbf{x}_{Ki} \in N_k(\mathbf{x}_{Kj})$, to reflect the structural information between the data, the weight matrix \mathbf{W}^K is constructed with the Gaussian kernel, and the construction of its elements as follows:

$$W_{ij}^k = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_{Ki} - \mathbf{x}_{Kj}\|^2}{t}\right), & \text{if } \mathbf{x}_{Ki} \in N_k(\mathbf{x}_{Kj}) \\ & \text{or } \mathbf{x}_{Kj} \in N_k(\mathbf{x}_{Ki}) \\ 0, & \text{other} \end{cases} \quad (19)$$

In Eq.19, $\|\mathbf{x}\| = (\sum_{i=1}^N \mathbf{x}_i^2)^{1/2}$ is the usual Euclidean (L2) norm, t is the Gaussian kernel parameter. \mathbf{W}^K models the local structure of the data manifold by determining whether the data points are neighbors and using the Gaussian kernel.

Fig.2 describes the decision hyperplanes of SVM and MCLPV_SVM on an artificial dataset. MCLPV_SVM fully

considers the connection between data samples, and thus it can be used to improve the large margin correlation filters.

B. MINIMUM CLASS LOCALITY PRESERVING VARIANCE CORRELATION FILTER

The SVM part of the large margin correlation filters uses the maximum margin of the boundary samples to establish the classification hyperplane, but it ignores the distribution information and internal relationships within the class, fails to fully utilize the data information, and the sample training is insufficient. Aiming at the above problem, inspired by the literature [33], we propose MCLPVCF as follows:

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \lambda \left(\mathbf{w}^T \mathbf{Z}_w \mathbf{w} + C \sum_{i=1}^N \xi_i \right) + (1 - \lambda) \sum_{i=1}^N \|\mathbf{w} \otimes \mathbf{x}_i - \mathbf{g}_i\|_2^2 \\ \text{s.t.} \quad & y_i (\mathbf{w}^T \mathbf{x}_i + b) \geq c_i - \xi_i, \xi_i \geq 0, i = 1, \dots, N \end{aligned} \quad (20)$$

In Eq.20, \mathbf{Z}_w is the locally preserving scatter matrix, and its definition is the same as that of Eq.17, Eq.18, and Eq.19, it uses the sample weighted adjacency graph to obtain the local manifold structure information. The parameters t and k can affect the value of \mathbf{Z}_w , t can change the specific value of the elements in the weight matrix, and k can determine the k sample points around the sample as neighbors, which influences the utilization of the manifold structure of the data.

C. THE SOLUTION OF MCLPVCF

Similar to the solution method in Section III, MCLPVCF is converted first to the frequency domain, then solved by the dual problem, and finally the filter is obtained by the SMO algorithm.

MCLPVCF in the frequency domain is as follows:

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \lambda \left(\widehat{\mathbf{w}}^\dagger \widehat{\mathbf{Z}}_w \widehat{\mathbf{w}} + C \sum_{i=1}^N \xi_i \right) + (1 - \lambda) \widehat{\mathbf{w}}^\dagger \widehat{\mathbf{Z}} \widehat{\mathbf{w}} \\ \text{s.t.} \quad & y_i (\widehat{\mathbf{w}}^\dagger \widehat{\mathbf{x}}_i + b') \geq c_i - \xi_i, \xi_i \geq 0, i = 1, \dots, N \end{aligned} \quad (21)$$

The dual problem of MCLPVCF is as follows:

$$\begin{aligned} \min_{\boldsymbol{\alpha}} \quad & \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \widehat{\mathbf{x}}_i^\dagger [\lambda \widehat{\mathbf{Z}}_w + (1 - \lambda) \widehat{\mathbf{Z}}] \widehat{\mathbf{x}}_j - \sum_{i=1}^N c_i \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^N \alpha_i y_i = 0; 0 \leq \alpha_i \leq \lambda C \end{aligned} \quad (22)$$

By solving the dual problem Eq.22 of MCLPVCF, the Lagrange multiplier vector $\boldsymbol{\alpha}^*$ is obtained, and $\widehat{\mathbf{w}}^*$ can be obtained using the following formula:

$$\widehat{\mathbf{w}}^* = [\lambda \widehat{\mathbf{Z}}_w + (1 - \lambda) \widehat{\mathbf{D}}]^{-1} \widehat{\mathbf{X}} \mathbf{Y} \boldsymbol{\alpha}^* \quad (23)$$

V. EXPERIMENTS

The experiments were divided into three parts. First, we briefly describe the process of the two algorithms. Second,

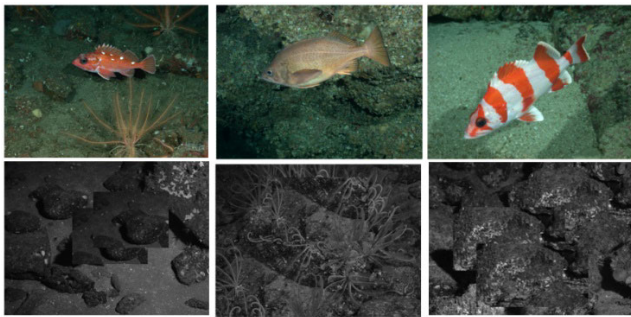


FIGURE 3. Partial images from the labeled fishes in the wild fish dataset.

we conducted different kinds of experiments to investigate the influence of these parameters. Finally, we report the results of the two algorithms in terms of target recognition and detection. The test environment was an Intel CPU i7 with 16 GB memory and MATLAB 2020a.

A. ALGORITHM PROCESS

Because the two proposed algorithms are both improved using large margin correlation filters, the algorithm steps are similar. First, the training sample images are normalized to obtain the training sample vectors \mathbf{x}_i , and Fourier transform is employed to convert \mathbf{x}_i to the frequency domain to obtain $\hat{\mathbf{x}}_i$. MCVLMCF uses $\hat{\mathbf{x}}_i$, y_i to calculate the within-class scatter matrix \hat{S}_ω , and MCLPVCF calculates the locally preserving scatter matrix \hat{Z}_w by Eq.17, Eq.18, and Eq.19. Then, calculate the matrix \hat{Z} , substitute the parameters into Eq.14 or Eq.22 and use SMO to obtain α^* . Finally, we substitute α^* and other parameters into Eq.15 or Eq.23 to solve the filter $\hat{\mathbf{w}}^*$. In the experiment, the filter was used to perform cross-correlation operations with the images to obtain the correlation output, and then judge the target position or category through the output. The algorithm process is given as follows.

B. PARAMETER INFLUENCE ON PERFORMANCE OF MCVLMCF

MCVLMCF has two parameters, C and λ , which have an influence on the performance. Therefore, we conducted a fish detection experiment to investigate the influence of these parameters on MCVLMCF. Fish detection is an important research topic in the field of ocean exploration. In this subsection, we test MCVLMCF with different parameters on the labeled fishes in the wild dataset of the NOAA fishery [34], as shown in Fig.3 is a partial image of the dataset. These images were collected through camera equipment deployed on remotely operated submersibles and belong to the dataset collected in the actual environment. We used grid search to find the optimal combination of parameters and some values can be determined empirically. The test results were also compared with other algorithms, including MMVCF [19], ZAMMCF [20], MOSSE [14], OTSDF [12], MCMF [16], and axisymmetric shell intersection-based correlation filter (ASICF) [35].

Algorithm 1 MCVLMCF and MCLPVCF

1. **Input:** Input N normalized images \mathbf{x}_i , and N_p is the number of positive training images, N_n is the number of negative training images.
2. **Convert to frequency domain:** Convert \mathbf{x}_i to frequency domain by Fourier transform.
3. **Solution of MCVLMCF:**
 - (1) for $i = 1: N_p$
Reshape positive training images \mathbf{x}_i to np column matrix \mathbf{x} and calculate its mean value.
end for
 - (2) for $i = 1: N_n$
Reshape positive training images \mathbf{x}_i to np column matrix \mathbf{x} and calculate its mean value.
end for
 - (3) Use the training image matrix of different classes \mathbf{x} and sample label y_i to calculate the within-class scatter matrix \hat{S}_ω by Eq.5.
 - (4) Construct the diagonal matrix \hat{X}_i , calculate the matrix \hat{Z} .
 - (5) Substitute the within-class scatter matrix \hat{S}_ω , the matrix \hat{Z} , the frequency domain image vector $\hat{\mathbf{x}}_i$, and parameters C , λ into Eq.14, and employ SMO to solve α^* .
 - (6) Substitute α^* , \hat{S}_ω , \hat{Z} into Eq.15 to obtain the frequency domain filter $\hat{\mathbf{w}}^*$.
4. **Solution of MCLPVCF:**
 - (1) for $i = 1: \in N_p$ to end do
if positive training images $\mathbf{x}_{Ki} N_k (\mathbf{x}_{Kj})$, then;
$$W_{ij}^k = \exp\left(-\|\mathbf{x}_{Ki} - \mathbf{x}_{Kj}\|^2 / t\right);$$

else
$$W_{ij}^k = 0;$$

end if
end for
 - for $i = 1: N_n$ to end do
if negative training images $\mathbf{x}_{Ki} N_k (\mathbf{x}_{Kj})$, then;
$$W_{ij}^k = \exp\left(-\|\mathbf{x}_{Ki} - \mathbf{x}_{Kj}\|^2 / t\right);$$

else
$$W_{ij}^k = 0;$$

end if
end for
 - Construct the matrix \mathbf{W}^K and \mathbf{D}^K using the elements W_{ij}^k and calculates the local preserving scatter matrix \hat{Z}_w by Eq.17 and Eq.18.
 - (2) Construct the diagonal matrix \hat{X}_i , calculate the matrix \hat{Z} .
 - (3) Substitute the locally preserving scatter matrix \hat{Z}_w , the matrix \hat{Z} , the frequency domain image vector $\hat{\mathbf{x}}_i$, and parameters C , λ into Eq.22, and employ SMO to solve α^*
 - (4) Substitute α^* , \hat{Z}_w , \hat{Z} into Eq.23 to obtain the frequency domain filter $\hat{\mathbf{w}}^*$
5. **Output:** Use the cross-correlation operation between the filter and the test image to obtain the correlation output. Obtain the localization result or the classification result by correlation output.

Because the fish in most of the images are not clear, this part of the images was excluded from the experiment.

TABLE 1. Accuracy of different parameters (%)

	$C=10^{-3}$	$C=10^{-2}$	$C=10^{-1}$	$C=10^0$	$C=10^1$	$C=10^2$
$\lambda=0.1$	92.5	93.7	93.7	93.7	93.7	93.7
	$D=0.2$	$D=0.2$	$D=0.2$	$D=0.2$	$D=0.2$	$D=0.2$
$\lambda=0.3$	82.5	82.5	82.5	82.5	82.5	82.5
	$D=0.3$	$D=0.3$	$D=0.3$	$D=0.3$	$D=0.3$	$D=0.3$
$\lambda=0.5$	83.7	83.7	83.7	83.7	83.7	83.7
	$D=0.3$	$D=0.3$	$D=0.3$	$D=0.3$	$D=0.3$	$D=0.3$
$\lambda=0.7$	83.7	83.7	83.7	83.7	83.7	83.7
	$D=0.3$	$D=0.3$	$D=0.3$	$D=0.3$	$D=0.3$	$D=0.3$
$\lambda=0.9$	57.5	57.5	57.5	57.5	57.5	57.5
	$D=0.4$	$D=0.4$	$D=0.4$	$D=0.4$	$D=0.4$	$D=0.4$
$\lambda=1$	36.2	36.2	36.2	38.7	30.0	30.0
	$D=0.4$	$D=0.4$	$D=0.4$	$D=0.4$	$D=0.4$	$D=0.4$

We randomly selected 80 images from the images containing clear fish as test samples, the remaining images containing clear fish were used as positive samples, and the clear images without fish were used as negative samples. During the training process, owing to the different sizes of the images, the training samples were changed into images of 160×160 pixels to obtain a fixed-size filter. In the test, we compared the test results with the ground-truth locations, and obtain the overlap rate with the ground-truth box to calculate the detection accuracy.

$$\begin{aligned}
 E_x &= \max(p_x + p_w, p'_x + p'_w), S_x = \min(p_x, p'_x) \\
 W &= p'_x + p'_w - (E_x - S_x) \\
 E_y &= \max(p_y + p_h, p'_y + p'_h), S_y = \min(p_y, p'_y) \\
 H &= p'_y + p'_h - (E_y - S_y) \\
 I &= W \times H
 \end{aligned} \tag{24}$$

In Eq.24, p_x, p_y, p'_x, p'_y are the horizontal and vertical coordinates of the upper left corner of the ground truth and the test detection box, respectively, p_w, p_h, p'_w, p'_h are the width and length of the ground truth and the test detection box, respectively, I is the overlapping area. The function used to calculate the overlap ratio O is as follows:

$$\begin{aligned}
 S_1 &= p_x \times p_y, S_2 = p'_x \times p'_y \\
 O &= \frac{I}{(S_1 + S_2 - I)}
 \end{aligned} \tag{25}$$

From Eq.24 and Eq.25 that the value of the overlap rate O must be in the range of $[0,1]$. O reflects the degree of overlap between the detection box and the real box (e.g., when $O=0$ the test detection box does not overlap with the ground truth). A threshold D is set and compared it with the overlap rate to determine whether the target is detected numerically.

Table 1 presents the detection accuracy of MCVLMCF for different parameters and thresholds. The test results show that MCVLMCF can achieve a particular detection accuracy for all parameter combinations. When the value of λ is 0.3 to 0.7, it has little influence on the detection results of MCVLMCF. When $\lambda=1$, that is, only the classification standard part is used, the detection accuracy decreases significantly.

Table 2 shows the detection accuracy of various algorithms in the fish detection experiment for different λ values at fixed

TABLE 2. Accuracy of different parameters λ (%).

λ	MMVCF	ZAMMCF	MMCF	MCVLMCF
0.1	52.5	52.5	50.0	81.2
0.2	51.3	52.5	50.0	81.2
0.3	50.0	51.3	48.8	82.5
0.4	48.8	51.3	47.5	82.5
0.5	53.8	53.8	51.3	83.7
0.6	55.0	56.3	51.3	83.7
0.7	57.5	57.5	53.8	83.7
0.8	58.8	58.8	57.5	82.5
0.9	55.0	57.5	55.0	81.2
1	36.3	35.0	35.0	61.2

TABLE 3. Accuracy of same threshold D .

D	MOSSE	OTSDF	MMVCF	ZAMMCF	ASICF	MMCF	MCVLMCF
0.3	36.3	48.8	58.8	58.8	81.2	57.5	82.5

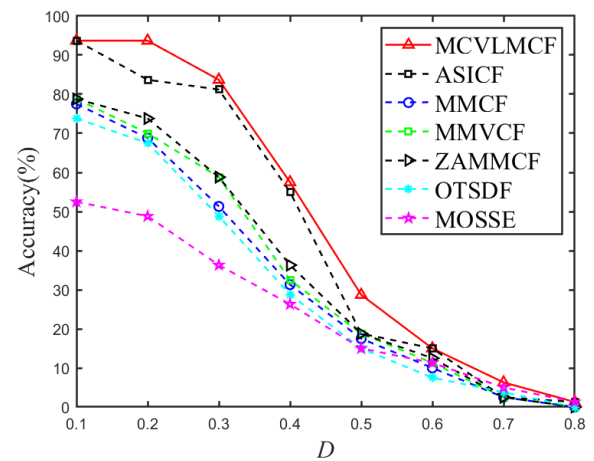


FIGURE 4. Detection accuracy of various algorithms on different values D .

$D=0.3, C=1$. Because only MMVCF, ZAMMCF, MMCF, and MCVLMCF have the parameter λ , the test results of these algorithms are compared. The accuracy of MCVLMCF is optimal for all set λ values, and it is also observed that MCVLMCF is not sensitive to variations in the value of λ .

Table 3 shows a comparison of the detection accuracy of various filters for fixed $D=0.3, \lambda=0.8, C=1$ (parameters set by MCVLMCF, MMVCF, ZAMMCF, and MMCF). In the fish detection experiment, MCVLMCF outperformed the other algorithms in terms of detection outcomes after applying the data distribution information.

The change curve of the detection accuracy for different methods under various D values with fixed parameters $C=1, \lambda=0.6$ is shown in Fig. 4. For all D values, MCVLMCF had a greater detection accuracy.

A fish detection example is shown in Fig. 5, where the red box represents the MCVLMCF detection box, the blue box represents the MMCF detection box, and the yellow box represents the ground truth box. It is evident that the MCVLMCF detection box has a larger overlap with the

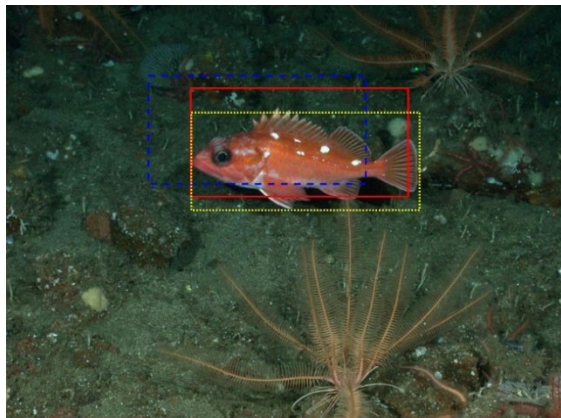
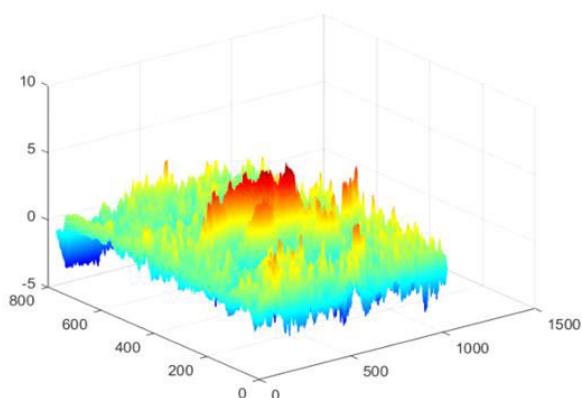
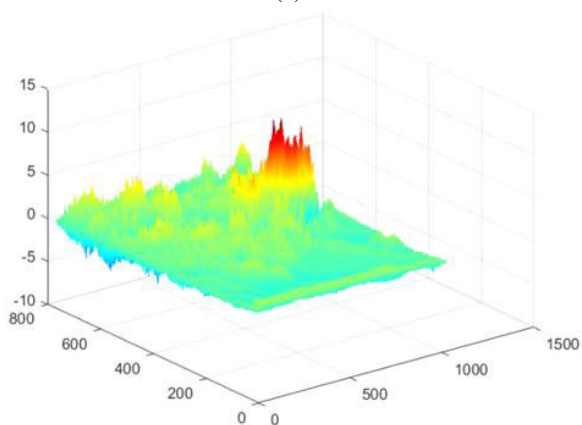


FIGURE 5. Detection boxes for MCVLMCF, MMCF, and ground truth.



(a)



(b)

FIGURE 6. Response output to the same image (a) MMCF (b) MCVLMCF.

ground truth box, indicating that the MCVLMCF detection result is more accurate.

The response of the identical detection image following filtering using MCVLMCF and MMCF is shown in Fig. 6. It is clear from a comparison of the two figures that MCVLMCF is better able to generate the peak response to the target with a higher response, more pronounced, and concentrated sharp

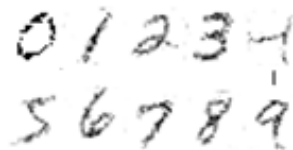


FIGURE 7. Partial image of MNIST dataset.

TABLE 4. Classification accuracy for different parameters k, C (%).

parameters	$C=10^{-3}$	$C=10^{-2}$	$C=10^{-1}$	$C=10^0$	$C=10^1$	$C=10^2$
$k=3$	94.5	98.5	97.5	96.5	96.5	96.5
$k=6$	94.5	95.0	98.7	96.5	96.5	96.5
$k=9$	94.0	95.0	98.7	97.0	97.0	97.0
$k=12$	94.0	95.0	98.5	97.0	97.0	97.0
$k=15$	94.0	94.5	98.5	97.5	97.0	97.0
$k=30$	93.5	94.5	96.5	98.5	97.0	97.0
$k=60$	93.0	94.5	95.5	98.7	97.0	97.0
$k=90$	93.0	94.0	95.0	98.7	97.0	97.0

fronts, whereas the other positions are flatter. The response values of MMCF are lower, the sharp fronts are less pronounced, and the range of fronts is wider.

The results in this subsection show that MCVLMCF performs better with the parameters $C=1$ and $\lambda=0.6$, and the parameters $C=1$ and $\lambda=0.6$ were adopted in the following experiments.

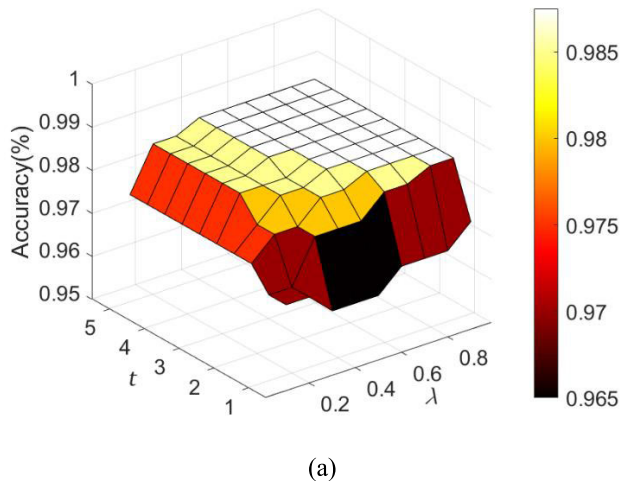
C. PARAMETER INFLUENCE ON PERFORMANCE OF MCLPVCF

We conducted a handwritten font classification experiment to examine the influence of MCLPVCF parameters, $C, \lambda, t,$ and k on the algorithm. We used the MNIST standard handwritten digit dataset [36] created by the National Institute of Standards and Technology for the experiment. A partial image of the MNIST dataset is shown in Fig.7. We randomly selected 400 images from each class as the training set and 600 as the test set, and combined the four parameters in pairs for testing.

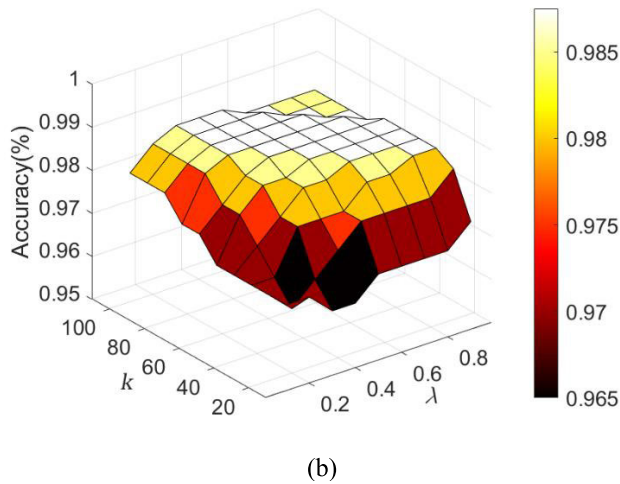
Fig.8 shows the classification accuracy of MCVPLCF for different parameters λ, t, k . The classification accuracy for fixed parameters $k=30, C=1$, the classification accuracy for fixed parameters $t=1, C=1$, and the classification accuracy for fixed parameters $\lambda=0.6, C=1$ are shown in Fig.8 (a), (b), and (c), respectively.

It can be found from Fig.8(a) that the classification accuracy is poor when t and λ are small, and the accuracy stabilizes when these two parameters increase to certain values. The situation in Fig.8(b) is roughly similar to that in Fig.8(a), but the accuracy drops instead when k and λ are large simultaneously. The accuracy diminishes as the value of k increases, as shown in Fig. 8(c), despite the relatively mild influence of k .

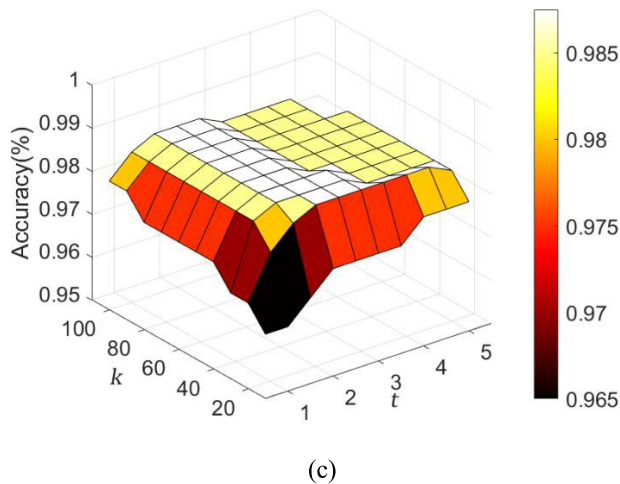
Referring to the parameter settings in the literature [33], Table 4 shows the classification accuracy for different parameters k and C , when the fixed parameters $\lambda=0.6$ and $t=1$.



(a)



(b)



(c)

FIGURE 8. Classification accuracy for different parameters (a) λ, t (b) λ, k (c) t, k .

Overall, the classification accuracy is higher when $k = 12, 15, 30$ and $C = 1$.

Table 5 shows the classification accuracy for different parameters t and C when the fixed parameter $\lambda = 0.6$ and $k = 30$. When $t = 1$ and 1.5 , the classification accuracy is

TABLE 5. Classification accuracy for different parameters t, C (%).

parameters	$C = 10^{-3}$	$C = 10^{-2}$	$C = 10^{-1}$	$C = 10^0$	$C = 10^1$	$C = 10^2$
$t = 0.5$	94.0	95.0	98.5	97.0	96.5	96.5
$t = 1.0$	93.5	94.5	96.5	98.5	97.0	97.0
$t = 1.5$	93.5	94.5	95.5	98.7	97.0	97.0
$t = 2.0$	93.0	94.5	95.5	98.7	97.0	97.0
$t = 2.5$	93.0	94.5	95.5	98.7	97.0	97.0

TABLE 6. Classification accuracy for different parameters λ, C (%).

parameters	$C = 10^{-3}$	$C = 10^{-2}$	$C = 10^{-1}$	$C = 10^0$	$C = 10^1$	$C = 10^2$
$\lambda = 0.1$	94.0	95.0	98.7	97.0	97.0	97.0
$\lambda = 0.3$	94.0	94.5	98.5	98.0	97.0	97.0
$\lambda = 0.5$	93.5	94.5	97.0	98.0	97.0	97.0
$\lambda = 0.7$	93.5	94.5	95.5	98.5	97.0	97.0
$\lambda = 0.9$	93.5	94.5	95.5	98.7	97.0	97.0
$\lambda = 1$	93.0	94.0	94.5	94.5	94.0	94.0



FIGURE 9. Yale dataset sample images.

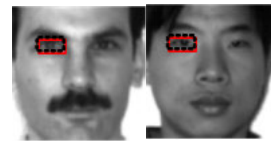


FIGURE 10. Examples of MCLPVCF and MCVMCF in eye localization experiment.

higher, and the influence of parameter t on the classification accuracy is substantially less than that of parameter k .

Table 6 shows the classification accuracy for different parameters λ and C , when the fixed parameter $t = 1$ and $k = 30$. It can be found that the classification accuracy is higher when λ is about 0.5 and 0.7 .

D. EYE LOCALIZATION

To compare the performance of the proposed algorithm with that of other algorithms, we conducted a human eye localization experiment. We used the Yale face dataset [37], created by Yale University, which contains 165 face images of 15 individuals, each category has 11 images with obvious expressions and lighting changes, and the size is 100×80 . A part of its image is shown in Fig.9. We randomly selected 45 images from it as the test set. In this experiment, the target was considered to be detected when the overlap rate $O \geq 0.3$.

Fig.10 shows two examples of MCLPVCF and MCVMCF in the eye localization experiment, in which the

TABLE 7. Localization accuracy in eye localization experiment.

MOSSE	MMVCF	ZAMMCF	MMCF	ASICF	MCVLMCF	MCLPVCF
80.0	88.9	91.1	88.9	91.1	93.3	93.3



FIGURE 11. COIL20 dataset sample images.

TABLE 8. Average classification accuracy (%) and standard deviation on the COIL20 dataset.

Algorithms	Average classification accuracy \pm standard deviation with different numbers of training images				
	20	25	30	35	40
MOSSE	76.9 \pm 1.7	77.1 \pm 0.9	78.0 \pm 0.7	78.5 \pm 0.7	78.7 \pm 0.5
OTSDF	90.9 \pm 1.4	91.2 \pm 0.8	92.7 \pm 0.7	93.9 \pm 0.8	94.4 \pm 0.4
MMVCF	90.9 \pm 1.6	92.0 \pm 0.9	92.9 \pm 0.7	93.8 \pm 0.8	94.0 \pm 0.7
ZAMMCF	91.1\pm1.7	92.0 \pm 0.8	93.1 \pm 0.6	94.2 \pm 0.9	94.3 \pm 0.9
MMCF	90.4 \pm 1.5	91.4 \pm 0.7	92.8 \pm 0.6	93.8 \pm 0.7	93.4 \pm 0.7
ASICF	90.5 \pm 1.6	92.0 \pm 0.4	92.5 \pm 0.7	93.4 \pm 0.7	94.0 \pm 0.5
AMSCF	90.6 \pm 1.5	92.1 \pm 0.8	92.7 \pm 0.8	94.2 \pm 0.4	94.2 \pm 0.4
MCVLMCF	90.6 \pm 1.6	92.2 \pm 0.7	93.5 \pm 0.7	94.4 \pm 0.7	94.4 \pm 0.5
MCLPVCF	90.8 \pm 1.5	93.2\pm0.8	94.4\pm0.8	95.8\pm0.7	95.7\pm0.6

red solid line box is obtained by MCLPVCF, and the black dotted box is obtained by MCVLMCF.

Table 7 shows the localization accuracy of MCLPVCF, MCVLMCF, and other algorithms. It can be found that the localization accuracies of MCLPVCF and MCVLMCF are at least 2.2% higher than those of the other algorithms.

E. CLASSIFICATION EXPERIMENT

In this subsection, we report the classification experimental results for multiple datasets. We contrasted the test results not only with the aforementioned methods but also with the adaptive manifold filter and spatial correlation feature (AMSCF) [38]. AMSCF is a method for hyperspectral image classification, but it can also process ordinary images. It also takes advantage of the data manifold structure.

In this study, seven datasets were tested. First, we employed three datasets for object classification experiments. The COIL20 dataset [39], created by Columbia University, contains 1440 grayscale images of 20 classes of objects, each class of objects are captured every 5 degrees, and each image is 128 \times 128 pixels. Fig.11 shows a partial sample image of this dataset. We randomly selected [20], [25], [30], [35], [40] images of each class as the training set and the remaining images as the test set. The test was repeated 40 times and the average value was used as the classification accuracy.

Table 8 lists the classification accuracies of different algorithms on the COIL20 dataset. MCLPVCF achieved most

TABLE 9. Average classification accuracy (%) and standard deviation on the COIL100 dataset.

Algorithms	Average classification accuracy \pm standard deviation with different numbers of training images				
	20	25	30	35	40
MOSSE	50.3 \pm 0.9	56.8 \pm 0.8	56.8 \pm 0.7	58.0 \pm 0.6	58.0 \pm 0.6
OTSDF	69.1 \pm 0.9	71.5 \pm 0.6	78.9 \pm 0.8	79.3 \pm 0.6	79.7 \pm 0.6
MMVCF	70.4 \pm 1.0	76.8 \pm 0.6	80.4 \pm 0.6	81.6 \pm 0.8	82.8 \pm 0.6
ZAMMCF	71.3 \pm 0.9	77.2 \pm 0.8	80.4 \pm 0.6	82.0 \pm 0.5	83.0 \pm 0.3
MMCF	70.4 \pm 0.8	76.4 \pm 0.6	80.0 \pm 0.7	80.0 \pm 0.6	82.2 \pm 0.6
ASICF	70.3 \pm 1.0	76.6 \pm 0.8	80.9 \pm 0.7	81.0 \pm 0.5	82.2 \pm 0.4
AMSCF	72.4 \pm 0.9	76.8 \pm 0.9	81.0 \pm 0.5	82.0 \pm 0.4	83.0 \pm 0.5
MCVLMCF	71.3 \pm 0.9	77.6 \pm 0.8	80.9 \pm 0.7	83.0 \pm 0.8	83.2 \pm 0.5
MCLPVCF	76.0\pm0.7	77.9\pm0.6	81.5\pm0.7	84.7\pm0.6	85.7\pm0.5



FIGURE 12. COIL100 dataset sample images.

TABLE 10. Average classification accuracy (%) and standard deviation on the Caltech101 dataset.

Algorithms	Average classification accuracy \pm standard deviation with different numbers of training images				
	20	25	30	35	40
MOSSE	49.3 \pm 0.9	50.8 \pm 0.5	51.2 \pm 0.6	51.9 \pm 0.7	52.4 \pm 0.5
OTSDF	66.7 \pm 0.9	67.3 \pm 0.5	68.5 \pm 0.7	69.3 \pm 0.7	69.5 \pm 0.6
MMVCF	69.4 \pm 1.0	70.5 \pm 0.5	71.4 \pm 0.8	71.8 \pm 0.8	72.5 \pm 0.6
ZAMMCF	68.8 \pm 0.8	72.0\pm0.8	72.0 \pm 0.6	72.5 \pm 0.6	72.8 \pm 0.5
MMCF	68.6 \pm 0.3	69.4 \pm 0.7	70.7 \pm 0.7	70.9 \pm 0.9	72.3 \pm 0.6
ASICF	68.8 \pm 0.3	69.6 \pm 0.7	70.5 \pm 0.7	70.5 \pm 0.9	72.3 \pm 0.6
AMSCF	71.0\pm1.2	71.2 \pm 0.6	71.2 \pm 0.5	73.1 \pm 0.8	73.5\pm0.8
MCVLMCF	71.0\pm1.0	71.5 \pm 0.8	71.9 \pm 0.6	73.2\pm0.9	73.4 \pm 0.6
MCLPVCF	70.5 \pm 0.8	71.6 \pm 0.3	72.5\pm0.5	73.1 \pm 0.6	73.5\pm0.5

of the optimal results, and MCLVCF basically achieved suboptimal results. Moreover, MCLPVCF is approximately 0.9%-17% and MCVLMCF is approximately 0.1%-15.7% high than other algorithms at training sample numbers of 25 to 40.

The COIL100 [39] dataset is shown in Fig.12, which is an extension of the COIL20 dataset, and contains 7200 color images of 100 classes of objects. Similarly, we randomly selected [20], [25], [30], [35], [40] images of each class as the training set and the remaining images as the test set. The test was repeated 40 times and the average value was used as the classification accuracy.

Table 9 shows the classification accuracies of the different algorithms on the COIL100 dataset. It can be found that there is an improvement of MCLPVCF on the COIL100 dataset compared to other algorithms. Moreover, MCVLMCF can achieve better classification accuracy. The COIL100 dataset has a large number of classes, and while objects seem similar in most images, the two proposed methods perform better



FIGURE 13. Caltech101 dataset sample images.

TABLE 11. Average classification accuracy (%) and standard deviation on the Yale face dataset.

Algorithms	Average classification accuracy \pm standard deviation with different numbers of training images				
	3	4	5	6	7
MOSSE	60.7 \pm 3.0	65.5 \pm 2.6	70.3 \pm 3.2	77.2 \pm 3.3	78.1 \pm 3.0
OTSDF	70.0 \pm 2.1	80.0 \pm 2.7	88.9 \pm 2.0	93.3 \pm 3.3	93.8 \pm 2.5
MMVCF	82.5 \pm 2.6	90.2 \pm 2.2	91.6 \pm 2.6	94.3 \pm 3.3	95.5 \pm 2.5
ZAMMCF	82.3 \pm 3.0	90.7 \pm 2.3	91.6 \pm 2.6	94.6 \pm 3.3	94.7 \pm 2.7
MMCF	81.7 \pm 3.0	89.5 \pm 2.6	90.0 \pm 2.6	93.3 \pm 2.0	94.4 \pm 2.0
ASICF	82.3 \pm 2.0	89.4 \pm 2.6	90.0 \pm 2.6	93.0 \pm 2.0	94.2 \pm 2.0
AMSCF	82.7 \pm 2.7	88.5 \pm 2.7	90.2 \pm 2.8	93.7 \pm 2.2	95.8 \pm 1.6
MCVLMCF	82.5 \pm 2.5	88.6 \pm 2.9	90.0 \pm 2.8	94.7 \pm 2.0	95.8 \pm 1.5
MCLPVCF	90.8 \pm 1.5	93.3 \pm 1.9	96.7 \pm 2.1	97.3 \pm 2.0	97.1 \pm 1.6



FIGURE 14. AR dataset sample images.

because they make full advantage of the spatial features of the data during training.

The Caltech101 dataset [40] is shown in Fig.13, which is created by the California Institute of Technology, and contains 9144 images of 101 categories. We randomly selected [20], [25], [30], [35], [40] images of each class as the training set and the remaining images as the test set.

Table 10 shows the classification accuracies of the different algorithms on the Caltech101 dataset. Most ideal and suboptimal outcomes are accomplished by MCVLMCF and MCLPVCF.

Then, we used four datasets for the face classification experiments. The Yale face dataset is shown in Fig.9, which is mentioned above. We randomly selected [3], [4], [5], [6], [7] images of each class as the training set and the remaining images as the test set. The test was repeated 30 times and the average value was used as the classification accuracy. Table 11 shows the classification accuracy of different algorithms.

The AR face dataset [41] shown in Fig.14 contains 3120 images of 120 people. The images have obvious



FIGURE 15. GT dataset sample images.

TABLE 12. Average classification accuracy (%) and standard deviation on the AR face dataset.

Algorithms	Average classification accuracy \pm standard deviation with different numbers of training images				
	5	10	15	17	20
MOSSE	63.4 \pm 2.4	70.9 \pm 2.5	71.7 \pm 1.7	72.7 \pm 1.8	72.3 \pm 2.0
OTSDF	80.5 \pm 2.0	82.5 \pm 1.9	80.4 \pm 1.9	80.7 \pm 1.8	80.8 \pm 1.7
MMVCF	92.6 \pm 2.0	93.4 \pm 1.7	93.5 \pm 2.2	94.7 \pm 1.8	95.8 \pm 2.1
ZAMMCF	90.6 \pm 2.0	93.6 \pm 1.8	95.7 \pm 2.0	95.8 \pm 1.7	96.7 \pm 1.6
MMCF	91.7 \pm 2.1	92.5 \pm 1.8	93.2 \pm 1.8	93.3 \pm 1.8	93.3 \pm 1.5
ASICF	91.0 \pm 2.2	92.0 \pm 1.8	93.1 \pm 1.8	93.2 \pm 1.6	93.5 \pm 1.5
AMSCF	89.8 \pm 2.3	93.2 \pm 1.9	96.2 \pm 1.9	96.4 \pm 1.5	97.0 \pm 1.6
MCVLMCF	89.5 \pm 2.3	93.8 \pm 1.9	96.2 \pm 1.8	96.7 \pm 1.8	97.2 \pm 1.5
MCLPVCF	90.7 \pm 2.0	97.5 \pm 1.7	98.6 \pm 1.6	98.7 \pm 1.6	99.2 \pm 1.5



FIGURE 16. ORL dataset sample images.

changes in lighting expressions, as well as obvious occlusion changes with glasses and scarves. Each image measured 40×50 pixels in size. We randomly selected [5], [10], [15], [17], [20] images of each class as the training set and the remaining images as the test set. The test was repeated 40 times and the average value was used as the classification accuracy.

Table 12 shows the classification accuracies of the different algorithms on the AR dataset. When the number of samples is large, the average classification accuracy of MCLPVCF and MCVLMCF is enhanced by 2.0%-27.3% and 0.2%-26.1% compared with the other algorithms, respectively.

The GT face dataset [42] shown in Fig. 15 contains 750 face images of 50 individuals. These images are color images with small changes in illumination, but large changes in pose and expression. We randomly selected [8], [9], [10], [11], [12] images of each class as the training set and the remaining images as the test set. The test was repeated 40 times and the average value was used as the classification accuracy.

TABLE 13. Average classification accuracy (%) and standard deviation on the GT face dataset.

Algorithms	Average classification accuracy \pm standard deviation with different numbers of training images				
	5	10	15	17	20
MOSSE	32.5 \pm 4.5	32.7 \pm 3.6	33.0 \pm 4.0	33.3 \pm 3.5	34.3 \pm 3.6
OTSDF	42.2 \pm 3.5	46.7 \pm 4.6	47.0 \pm 4.0	51.3 \pm 2.5	52.3 \pm 2.7
MMVCF	58.2 \pm 3.6	58.6 \pm 4.0	60.8 \pm 3.8	63.3 \pm 3.3	66.7 \pm 3.0
ZAMMCF	58.7\pm4.7	58.9 \pm 3.8	59.6 \pm 4.0	64.3 \pm 4.3	64.6 \pm 4.6
MMCF	57.9 \pm 4.1	57.5 \pm 4.0	59.0 \pm 4.0	62.5 \pm 4.5	63.5 \pm 4.5
ASICF	57.8 \pm 4.0	57.4 \pm 4.0	59.5 \pm 4.2	62.3 \pm 4.5	63.5 \pm 4.6
AMSCF	58.0 \pm 4.5	57.5 \pm 4.3	60.5 \pm 4.2	64.0 \pm 4.3	66.7 \pm 4.6
MCVLMCF	57.9 \pm 4.2	57.5 \pm 4.1	60.8 \pm 4.0	64.8 \pm 4.3	66.7 \pm 4.5
MCLPVCF	55.7 \pm 4.0	61.7\pm4.0	67.0\pm4.0	68.8\pm4.3	68.4\pm4.0

TABLE 14. Average classification accuracy (%) and standard deviation on the ORL face dataset.

Algorithms	Average classification accuracy \pm standard deviation with different numbers of training images				
	5	10	15	17	20
MOSSE	54.2 \pm 1.5	63.3 \pm 1.8	75.0 \pm 1.5	74.3 \pm 1.5	75.0 \pm 1.6
OTSDF	75.7 \pm 1.7	83.3 \pm 1.6	91.0 \pm 1.5	91.3 \pm 1.5	95.0 \pm 1.7
MMVCF	86.0 \pm 1.0	86.6 \pm 1.6	91.6 \pm 1.6	92.4 \pm 1.6	95.5 \pm 1.5
ZAMMCF	86.8 \pm 1.7	87.7 \pm 1.7	91.0 \pm 2.0	93.7 \pm 1.5	95.0 \pm 1.6
MMCF	85.8 \pm 2.1	85.6 \pm 1.3	90.0 \pm 2.0	92.5 \pm 1.0	95.0 \pm 1.5
ASICF	85.6 \pm 2.0	85.8 \pm 1.5	90.0 \pm 2.2	92.3 \pm 1.2	94.0 \pm 1.6
AMSCF	89.1 \pm 2.1	91.0 \pm 2.0	92.0 \pm 1.6	94.0 \pm 2.0	95.7 \pm 1.5
MCVLMCF	89.2 \pm 2.0	90.0 \pm 2.1	95.0\pm1.0	95.0 \pm 2.0	96.7\pm1.5
MCLPVCF	91.7\pm1.7	91.4\pm2.1	95.0\pm1.5	96.3\pm1.3	96.7\pm1.7

Table 13 lists the classification accuracies of the different algorithms on the GT dataset. It can be found that the average classification accuracy of all algorithms is not high, which is due to the large variation in face pose, expression, and illumination in this dataset. However, MCLPVCF and MCVLMCF still achieve the optimal and suboptimal classification accuracies with larger number of training samples.

The ORL face dataset [43], created by the University of Cambridge, contains 400 images of 50 individuals with obvious illumination and pose changes. Each image has a pixel size of 92×112 . A part of the image is shown in Fig.16. We randomly selected [3], [4], [5], [6], [7] images of each class as the training set and the remaining images as the test set. The test was repeated 40 times and the average value was used as the classification accuracy.

Table 14 shows the classification accuracies of the different algorithms on the ORL dataset. MCLPVCF and MCVLMCF are 0.4%-37.5% and 0.1%-35% higher than other algorithms, respectively.

We also used the ANOVA test to verify whether the accuracies obtained in the above datasets are credible, and the results obtained are significantly less than 0.05. Therefore, we believe that the results obtained are plausible and our proposed algorithms make a significant difference in the accuracy of the different methods. In general, it can be found that in the object classification and face classification experiments, MCLPVCF has achieved most of the best results,

TABLE 15. Experimental time on the different dataset.

Algorithms(dataset)	Training time(s)	Test time(s)
MCVLMCF (Yale)	3.985	0.35
MCLPVCF (Yale)	4.836	0.35
MMCF (Yale)	2.464	0.35
MCVLMCF (GT)	9.886	0.31
MCLPVCF (GT)	10.547	0.31
MMCF (GT)	7.191	0.31
MCVLMCF (ORL)	6.478	0.35
MCLPVCF (ORL)	7.652	0.35
MMCF (ORL)	4.327	0.35

MCVLMCF has achieved most of the sub-optimal and some of the best accuracies. Thus, the proposed methods outperformed the other algorithms in terms of the classification performance.

F. COMPUTATIONAL COMPLEXITY

The training time and test time of the two proposed algorithms and MMCF on the Yale, GT, and ORL face datasets are presented in Table 15. The experimental parameters were set to 105 training images and 60 test images of the Yale dataset, and the image dimension was $d = 60 \times 60$; 600 training images, 150 test images of the GT dataset, and the image dimension was $d = 50 \times 60$; 280 training images, 120 test images of the ORL dataset, and the image dimension was $d = 60 \times 60$. The training time in the table is the time used to train a filter template, and the test time is the average time per image. The three methods below solve the QP problem and Fourier transform, which accounts for the majority of their computational costs. However, the two proposed methods require additional computations to compute the within-class scatter matrix for MCVLMCF and the locally preserving scatter matrix for MCLPVCF, which adds to their training time. But their testing time is the same. We believe that it is worthwhile to invest slightly more time in training to achieve greater performance.

VI. CONCLUSION

In this paper, we have described the problems with large margin correlation filters, such as not sufficiently considering the sample distribution and not applying the sample internal structure. To address these problems, we propose two methods, MCVLMCF and MCLPVCF. First, MCVLMCF introduces the within-class scatter into the correlation filter, which takes into account the class distribution of the data while using the principle of large margin, and fully utilizes the sample dispersion and its data information so that the data of the same class can be more compact. At the same time, the excellent generalization performance of the SVM and the localization ability of the correlation filter are used to make the obtained response more in line with the real situation. Then, through the study of weighted adjacency graph, MCLPVCF utilizes the structural information between samples, introduces the locally preserving scatter

information of the data, takes the distribution information and intrinsic manifold structure of the samples into full consideration during training, and maximizes the classification interval and optimizes the correlation output. The construction of the filter takes into account the category information and structural information of the sample, which can also improve the detection and classification performance of the algorithm. The experimental results show that MCVLMCF and MCLPVCF perform well in localization and classification experiments. The future work is to optimize the time complexity of MCVLMCF and MCLPVCF to overcome the expensive training time cost. Additionally, we will verify the feasibility of introducing other variants of SVM into the large margin correlation filters.

REFERENCES

- [1] M. Boyuk, R. Duvar, and O. Urhan, "Deep learning based vehicle detection with images taken from unmanned air vehicle," in *Proc. Innov. Intell. Syst. Appl. Conf. (ASYU)*, Oct. 2020, pp. 1–4, doi: [10.1109/ASYU50717.2020.9259868](https://doi.org/10.1109/ASYU50717.2020.9259868).
- [2] K. S. Wickramasinghe and G. U. Ganegoda, "Pedestrian detection, tracking, counting, waiting time calculation and trajectory detection for pedestrian crossings traffic light systems," in *Proc. 20th Int. Conf. Adv. ICT Emerg. Regions (ICTer)*, Nov. 2020, pp. 172–177, doi: [10.1109/ICTer51097.2020.9325479](https://doi.org/10.1109/ICTer51097.2020.9325479).
- [3] S. Namiki, K. Yokoyama, S. Yachida, T. Shibata, H. Miyano, and M. Ishikawa, "Online object recognition using CNN-based algorithm on high-speed camera imaging: Framework for fast and robust high-speed camera object recognition based on population data cleansing and data ensemble," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 2025–2032, doi: [10.1109/ICPR48806.2021.9413042](https://doi.org/10.1109/ICPR48806.2021.9413042).
- [4] R. He, J. Cao, L. Song, Z. Sun, and T. Tan, "Adversarial cross-spectral face completion for NIR-VIS face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 5, pp. 1025–1037, May 2020, doi: [10.1109/TPAMI.2019.2961900](https://doi.org/10.1109/TPAMI.2019.2961900).
- [5] H. Tao, X. Shen, and Q. Deng, "Infrared target tracking algorithm based on Bernoulli filter and support vector machine," in *Proc. Int. Conf. Inf. Sci. Educ. (ICISE-IE)*, Dec. 2020, pp. 277–281, doi: [10.1109/ICISE51755.2020.00067](https://doi.org/10.1109/ICISE51755.2020.00067).
- [6] A. V. Lugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inf. Theory*, vol. IT-10, no. 2, pp. 139–145, Apr. 1964, doi: [10.1109/TIT.1964.1053650](https://doi.org/10.1109/TIT.1964.1053650).
- [7] C. F. Hester and D. Casasent, "Multivariate technique for multiclass pattern recognition," *Appl. Opt.*, vol. 19, no. 11, pp. 1758–1761, 1980.
- [8] B. V. K. V. Kumar, "Minimum-variance synthetic discriminant functions," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 3, no. 10, pp. 1579–1584, 1986.
- [9] A. Mahalanobis, B. V. K. V. Kumar, and D. Casasent, "Minimum average correlation energy filters," *Appl. Opt.*, vol. 26, no. 17, pp. 3633–3640, 1987.
- [10] A. Mahalanobis, B. V. K. V. Kumar, S. Song, S. R. F. Sims, and J. F. Epperson, "Unconstrained correlation filters," *Appl. Opt.*, vol. 33, no. 17, pp. 3751–3759, 1994.
- [11] B. Vijaya Kumar, "Minimum squared error synthetic discriminant functions," *Opt. Eng.*, vol. 31, no. 5, pp. 915–931, 1992.
- [12] B. V. K. V. Kumar, A. Mahalanobis, and D. W. Carlson, "Optimal trade-off synthetic discriminant function filters for arbitrary devices," *Opt. Lett.*, vol. 19, no. 19, pp. 1556–1558, 1994.
- [13] D. S. Bolme, B. A. Draper, and J. R. Beveridge, "Average of synthetic exact filters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2105–2112, doi: [10.1109/CVPR.2009.5206701](https://doi.org/10.1109/CVPR.2009.5206701).
- [14] D. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2544–2550, doi: [10.1109/CVPR.2010.5539960](https://doi.org/10.1109/CVPR.2010.5539960).
- [15] A. Rodriguez, V. N. Boddeti, B. V. K. V. Kumar, and A. Mahalanobis, "Maximum margin correlation filter: A new approach for localization and classification," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 631–643, Feb. 2013, doi: [10.1109/TIP.2012.2220151](https://doi.org/10.1109/TIP.2012.2220151).
- [16] J. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA, USA: MIT Press, 1999, pp. 185–208.
- [17] Y. Wang and Q. Wu, "Research on face recognition technology based on PCA and SVM," in *Proc. 7th Int. Conf. Big Data Anal. (ICBDA)*, Mar. 2022, pp. 248–252, doi: [10.1109/ICBDA55095.2022.9760320](https://doi.org/10.1109/ICBDA55095.2022.9760320).
- [18] Z. Yi and H. Xiaoyong, "Research on pedestrian detection system based on tripartite fusion of 'HOG+SVM+median filter,'" in *Proc. Int. Conf. Artif. Intell. Comput. Eng. (ICAICE)*, Oct. 2020, pp. 484–488, doi: [10.1109/ICAICE51518.2020.00100](https://doi.org/10.1109/ICAICE51518.2020.00100).
- [19] V. Naresh Boddeti and B. V. K. V. Kumar, "Maximum margin vector correlation filter," 2014, *arXiv:1404.6031*.
- [20] J. A. Fernandez, V. N. Boddeti, A. Rodriguez, and B. V. K. V. Kumar, "Zero-aliasing correlation filters for object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 8, pp. 1702–1715, Aug. 2015, doi: [10.1109/TPAMI.2014.2375215](https://doi.org/10.1109/TPAMI.2014.2375215).
- [21] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 831–836, Aug. 1996, doi: [10.1109/34.531802](https://doi.org/10.1109/34.531802).
- [22] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997, doi: [10.1109/34.598228](https://doi.org/10.1109/34.598228).
- [23] X. He, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16, 2003, pp. 186–197.
- [24] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005, doi: [10.1109/TPAMI.2005.55](https://doi.org/10.1109/TPAMI.2005.55).
- [25] E. Kokiopoulou and Y. Saad, "Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2143–2156, Dec. 2007, doi: [10.1109/TPAMI.2007.1131](https://doi.org/10.1109/TPAMI.2007.1131).
- [26] V. N. Boddeti, T. Kanade, and B. V. K. V. Kumar, "Correlation filters for object alignment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2291–2298, doi: [10.1109/CVPR.2013.297](https://doi.org/10.1109/CVPR.2013.297).
- [27] J. Inglada, "Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features," *ISPRS J. Photogramm. Remote Sens.*, vol. 62, no. 3, pp. 236–248, Aug. 2007.
- [28] S. Zafeiriou, A. Tefas, and I. Pitas, "Minimum class variance support vector machines," *IEEE Trans. Image Process.*, vol. 16, no. 10, pp. 2551–2564, Oct. 2007, doi: [10.1109/TIP.2007.904408](https://doi.org/10.1109/TIP.2007.904408).
- [29] X. Li, X. Jia, L. Wang, and K. Zhao, "Reduction of spectral unmixing uncertainty using minimum-class-variance support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1335–1339, Sep. 2016, doi: [10.1109/LGRS.2016.2584660](https://doi.org/10.1109/LGRS.2016.2584660).
- [30] R. Chen, J. Hu, and M. R. Asharif, "Adaptive filter by using proportionate frequency domain extended correlation LMS algorithm in the double-talk condition," in *Proc. 9th Int. Conf. Signal Process.*, Oct. 2008, pp. 1826–1829, doi: [10.1109/ICOSP.2008.4697495](https://doi.org/10.1109/ICOSP.2008.4697495).
- [31] A. V. Oppenheim, A. S. Willksy, and S. Hamid, *Signals and Systems*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1997.
- [32] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," in *Advances in Kernel Methods: Support Vector Learning*, vol. 208, no. 14. Cambridge, MA, USA: MIT Press, 1998, pp. 98–112.
- [33] X. Wang, F.-L. Chung, and S. Wang, "On minimum class locality preserving variance support vector machine," *Pattern Recognit.*, vol. 43, no. 8, pp. 2753–2762, Aug. 2010.
- [34] G. Cutter, K. Stierhoff, and J. Zeng, "Automated detection of rockfish in unconstrained underwater videos using Haar cascades and a new image dataset: Labeled fishes in the wild," in *Proc. IEEE Winter Appl. Comput. Vis. Workshops*, Jan. 2015, pp. 57–62, doi: [10.1109/WACVW.2015.11](https://doi.org/10.1109/WACVW.2015.11).
- [35] Z. Xie, Y. Zhang, S. Chi, L. Zhou, and M. Li, "Adaptive target detection algorithm based on correlation filtering," in *Proc. IEEE Int. Conf. Unmanned Syst. (ICUS)*, Oct. 2019, pp. 124–130, doi: [10.1109/ICUS48101.2019.8996036](https://doi.org/10.1109/ICUS48101.2019.8996036).
- [36] Y. Lecun. (Aug. 2019). *The MNIST Database of Handwritten Digits*. [Online]. Available: <http://yann.lecun.com/exdb/mnist/>
- [37] A. S. Georgeiades, P. N. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001, doi: [10.1109/34.927464](https://doi.org/10.1109/34.927464).

- [38] J. Liao and L. Wang, "Adaptive hyperspectral image classification based on the fusion of manifolds filter and spatial correlation features," *IEEE Access*, vol. 8, pp. 90390–90409, 2020, doi: [10.1109/ACCESS.2020.2993864](https://doi.org/10.1109/ACCESS.2020.2993864).
- [39] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *Int. J. Comput. Vis.*, vol. 14, no. 1, pp. 5–24, Jan. 1995.
- [40] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop*, 2004, p. 178, doi: [10.1109/CVPR.2004.383](https://doi.org/10.1109/CVPR.2004.383).
- [41] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based Fisher discrimination dictionary learning for image classification," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 209–232, Sep. 2014.
- [42] A. V. Nefian and M. H. Hayes, "A hidden Markov model-based approach for face detection and recognition," Ph.D. dissertation, Georgia Inst. Technol., Atlanta, GA, USA, 1999.
- [43] F. S. Samaria and A. C. Harter, "Parameterization of a stochastic model for human face identification," in *Proc. IEEE Workshop Appl. Comput. Vis.*, vol. 22, 1994, pp. 138–142.



JIA LUO received the M.E. degree in computer software and theory from Sichuan Normal University, Chengdu, China. She is currently a Professor with the School of Computer Science, Sichuan University Jinjiang College, Meishan. Her current research interests include image processing and pattern recognition.



GANG ZHOU received the M.E. degree in computer software and theory from the University of Electronic Science and Technology of China, Chengdu, China. He is currently a Teacher with the Department of Data Science and Big Data Technology, Sichuan University Jinjiang College, Meishan. His current research interests include machine learning, big data, and the IoT applications.



SAMYA BAGCHI was born in Calcutta, India. He received the B.Tech. degree in information technology from WBUT, in 2011, and the M.Tech. degree in information technology from IIITB, in 2013. He is currently pursuing the Ph.D. degree with the Australian Institute of Machine Learning, University of Adelaide. He started his career as a Research Engineer at Siemens Corporate Research in the domain of parallel computing, where he got interested in artificial intelligence and neuro-morphic computing. With the vision to apply AI to improve healthcare, he started working on his start-up through which he got selected as an Ignite Scholar at the Stanford Graduate School of Business. He has also worked as an AI-Consultant, a YC-backed start-up CureSkin. He is also working on satellite star tracking with the University of Adelaide. He is also the Founder of Anixr, where the focus is to use AI to reduce animation rendering cost. He is also actively working with an investment banker to use AI in Forex trading.



QI JIANG received the M.E. degree in computer science and technology from Xihua University, Chengdu, China. He is currently a Teacher with the Department of Computer Science and Technology, Sichuan University Jinjiang College, Meishan. His current research interests include machine learning, pattern recognition, and image classification.

...