SHORT PAPER

Feature fusion based on joint sparse representations and wavelets for multiview classification

Younes Akbari[1] · Omar Elharrouss[1] · Somaya Al-Maadeed[1]

Abstract

Feature-level-based fusion has attracted much interest. Generally, a dataset can be created in different views, features, or modalities. To improve the classification rate, local information is shared among different views by various fusion methods. However, almost all the methods use the views without considering their common aspects. In this paper, wavelet transform is considered to extract high and low frequencies of the views as common aspects to improve the classification rate. The fusion method for the decomposed parts is based on joint sparse representation in which a number of scenarios can be considered. The presented approach is tested on three datasets. The results obtained by this method prove competitive performance in terms of the datasets compared to the state-of-the-art results.

Keywords Fusion method · Feature extraction · Wavelet transform

1 Introduction

Due to the complexity of some tasks in different research areas [7, 15], even when using deep learning techniques, exploiting different features using fusion approaches to obtain the final results can provide some improvements for the tasks [14]. It is well explored that the fusion process can improve the classification rate when multiple sources are considered [20]. The process aims to provide a combination of local information in different views [44]. For classification problems, fusion can occur at the feature [37] and classifier levels [38]. Fusion at the feature level aggregates features extracted from multiple sources into a common space that can be represented as a single space or separated spaces. Fusion at the classifier level obtains a decision from a combination of individual classifiers by training each view separately. Feature fusion was developed from concatenating features simply to complex fusion methods. Although a new fusion method can obtain better

results than a traditional method (concatenating features), existing noise among the features can affect their accuracy. To solve this problem, different approaches have been used. One of the approaches concentrates on separating the views into common aspects [21].

To solve these problems, this paper presents a novel fusion approach in which the views (feature extraction methods) are separated into high and low frequencies using the wavelet transform. The wavelet transform separates the components of the features into different frequency bands, allowing a sparser representation of the features. Simultaneously, the decomposition reduces the impact of noise on the fusion methods. The values of the wavelet coefficient with respect to the low- and high-frequency subbands disclose important information related to the signal structure (feature). Both frequency subbands usually imply spikes, with the high and low values of the wavelet coefficient corresponding to complex spikes and a smooth region, respectively. Since the effects of the separated information based on the wavelet transform were presented in [4, 5, 6, 40], the separated information, which can be any change in the original feature, such as applying filters to the signal, is useful for improving the accuracy of the classification step. The purpose of using low- and high-frequency wavelet subbands is to filter out the noise while preserving the feature map structures very well. The noise in the features affects the features learned by the classifier. We also want to preserve the structures of the feature maps. These structures

✉ Younes Akbari
  Younes.akbari@qu.edu.qa

  Omar Elharrouss
  elharrouss.omar@gmail.com

  Somaya Al-Maadeed
  s_alali@qu.edu.qa

1 Department of Computer Science and Engineering, Qatar University, Doha, Qatar

can be very useful for feature-level fusion tasks. We use joint sparse representation for fusing low and high frequencies as one of the popular fusion methods. Figure 1 illustrates the overview of our proposed approach. As mentioned above, the fusion methods, especially sparse representation methods, suffer noise of features, which can affect their results [16]. Therefore, to avoid this problem, wavelets can be a sufficient choice. Additionally, we explore the impact of decomposition levels and two states that can occur in the fusion step. The first state is shown in Fig. 1, namely low and high frequencies are fed into the fusion method separately, and the other state, with all of the frequencies, can be fused simultaneously. The main contributions of this work are described as follows:

– *Multifeature fusion approach:* A novel fusion approach combining wavelets and joint sparse representation is presented.
– *Exploring view separation in a fusion approach*: Separating features in a sparse frequency space for classification problems is investigated for the first time (to the best of the authors' knowledge).

– *Improved accuracy for multiview classification*: We show that compared with the fusion approaches, the proposed methods achieve superior performance.

The remainder of this paper is organized as follows. Section 2 provides an overview of the related works, and the proposed approach is presented in Sect. 3. Experimental results are reported in Sect. 4, while Sect. 5 concludes this paper.

2 Related works

The aim of the multifeature is to reveal and relate the correlation of features across different views. Approaches to address the aim (similarity across features) can be categorized into three groups: multikernel learning [29, 39], subspace learning [22, 46], and sparse representation [1, 2, 8]. Since we focus on the sparse representation approach, we explore the state-of-the-art category. Due to the attractiveness of many researchers in using sparse representation, approximating data by considering a few dictionary atoms
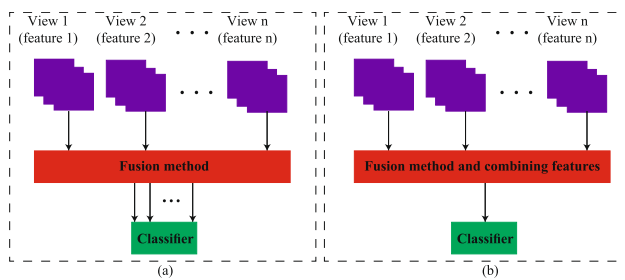
Fig. 1 Overview of proposed system
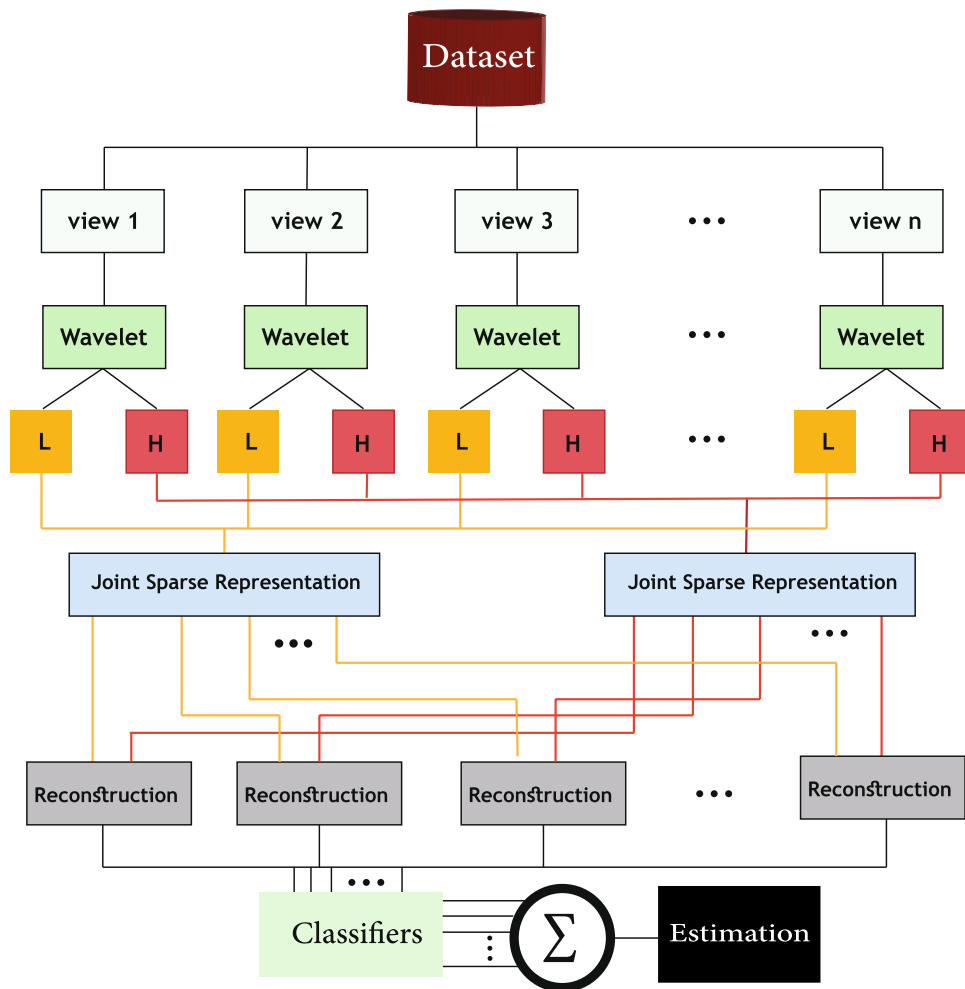
Fig. 2 Different structures for multifeature fusion

was proposed [1, 8, 19, 25, 26, 27, 28, 30, 31, 49, 50, 52]. A relaxed collaborative representation (RCR) approach was proposed in [49]. They assumed to represent different features that can consider their coefficients common. This obtains a result by minimizing the sparse codes through counting the sum of the distance of coefficients from their average. Reference [50] considered the l_1, l_2 norm to obtain a joint sparse representation for the multiple features (MTJSRC) and tested their methods on the data with high-dimensionality. Li et al. [27] proposed a joint discriminative collaborative representation (JDCR) approach to fuse multiple features with the aim of obtaining both similarities and discriminatively the representation coefficients. Reference [19] presented a joint feature extraction to align multifeature groups and introduced a feature selection method for dimensionality reduction. Partial multiview clustering (PVC) was presented in [30] in which data were considered incomplete. They used nonnegative matrix factorization (NMF) [25] to train a latent subspace. In [8] and [26], a sparse representation model based on dictionary learning was introduced that obtained promising results when multimodal features were considered. Due to the assumption of missing data in the multifeature extraction step, Zhao et al. [52] presented a partial multifeature unsupervised framework by preserving the similarity structure across different features. Nonparametric sparsity-based learning to reduce the dimensionality of multiple features using the matrix decomposition method was presented in [31]. In [28], to learn multiple features extracted for the problem of diabetes mellitus and impaired glucose regulation, both specific and similar components were used, and effective results were reported.

Although the mentioned methods to fuse multiple features have achieved promising results in different classification and clustering applications, the methods can be improved by some changes. Thus, we present a novel multiview learning approach to improve the methods. In general, more methods use all features simultaneously and follow two common structures as shown in Fig. 2. In the first structure (Fig. 2a), the fusion method is applied to all

views, and the result is a set of fused views corresponding to each view, i.e. the number of views after fusion is equal to the original views. Then, a classifier can be used for each view. Finally, fusion can be performed with the classifiers. In the second structure (Fig. 2b), the output of the fusion method is a single feature space that can be fed into a classifier. The method proposed in this study uses the first structure.

3 Proposed method

Background information about our steps, including wavelet transform and joint sparse representation, and our implementation are presented in the following subsections.

3.1 Wavelet transform

Wavelets as a tool can analyse signals with discontinuities and sharp spikes. To implement this tool, the high-pass and low-pass functions are exploited. Structures such as high and fast fluctuations are better preserved and can be used for noise removal when compared to other transforms. Additionally, they can play a great role in extracting features. To implement a 1D wavelet transform, we use two high-pass and low-pass filters [34]. Let $S = \{s_1, s_2, ..., s_N\}$ be the dataset captured in N different views (feature extraction methods). For each view, we have a feature vector whose length varies based on the feature extraction method. We compute approximation (cA) and detail (cD) subbands for the decomposition step by convolving the set with a high-pass filter (Hi_F) for obtaining detail coefficients and a low-pass filter (Lo_F) for obtaining approximation coefficients:

$$cD = s_n * Hi\_F(Hwavelet), \tag{1}$$

$$cA = s_n * Lo\_F(Lwavelet), \tag{2}$$

where the H and L wavelets correspond to high frequency and low frequency, respectively. Additionally, they are presented as the low-pass and high-pass function results.

3.2 Multiview join sparse representation

Since wavelets produce frequencies in sparse space, we select a fusion method based on sparse representation. An efficient tool for fusing multiple features is joint sparse representation [12, 51]. If we have $FE = [1, ..., FE]$ as a finite set of available feature extraction methods and $X^{FE} = [x_1^{fe}, x_2^{fe}, ..., x_N^{fe}] \in \mathbb{R}^{n^{fe} \times N}, fe \in FE$ as the collection of N (normalized) training samples of the methods, we can statistically assume the independence of the data (x^{fe} is the

feature vector for the sth method (view)). To address the fusion step, the method formulates it by dictionary representation $D^{fe} \in \mathbb{R}^{nfe \times d}$ corresponding to the sth method. Therefore, we have the multi-feature dictionaries constructed by data extracted from different methods. That is, the jth atom of dictionary D^{fe} is the jth data produced by the fe^{th} method. If $\left\{ x^{fe} \mid fe \in FE \right\}$ is a multifeature sample, then we can solve the ι_{12}-regularized reconstruction problem to obtain the optimal code sparse matrix $A^* \in \mathbb{R}^{d \times FE}$:

$$l(x, D) \doteq \min_{A\left[\alpha^1 \dots \alpha^{FE}\right]} \frac{1}{2} \sum_{fe=1}^{FE} \left\| x^{fe} - D^{fe}\alpha^{fe} \right\|_{\iota_2}^2$$
$$+ \lambda_1 \|A\|_{\iota_{12}} + \frac{\lambda_2}{2} \|A\|_F^2, \tag{3}$$

where the regularizing parameters are λ_1 and λ_2. To obtain a unique solution, the Frobenius norm $\|.\|_F$ term is added for the joint sparse optimization problem [8]. Here, α^{fe} is the fe^{th}- column of A which shows the sparse representation for the fe^{th} method. The ι_2 norm of a vector $x \in \mathbb{R}^m$ and the ι_{12} norm of matrix $X \in \mathbb{R}^{m \times n}$ are defined as $\|x\|_{\iota_2} = (\sum_{j=1}^{m} |x_j|^2)^{1/2}$ and $\|X\|\iota_{12} = \sum_{i=1}^{m} \|x_{i\rightarrow}\|_{\iota_2}$ ($x_{i\rightarrow}$ is the ith row of matrix), respectively. To solve the optimization problem, several algorithms have been proposed [36] wherein to find A^*, we applied the efficient method of multipliers (ADMM) [35]. Multimodal dictionaries are obtained by the optimization problem:

$$D^{fe*} = \arg\min_{D^{fe} \in \mathbb{D}} E_{x^{fe}}[l(x^{fe}, D^{fe})], \quad \forall fe \in FE \tag{4}$$

where the convex set $\mathbb{D}$ is defined as:

$$\mathbb{D}^{fe} \doteq \left\{ D \in \mathbb{R}^{nfe \times d} \mid \|d_k\|_{\iota_2} \le 1, \ \forall k = 1, ..., d \right\}. \tag{5}$$

Data x^{fe} are assumed to come from a finite (unknown) probability distribution $p(x^{fe})$. A classical projected stochastic gradient algorithm [3] can be used to solve the optimization problem above and gives a sequence of updates for each iteration:

$$D^{fe} \leftarrow \Pi_{\mathbb{D}^{fe}}[D^{fe} - \rho_t \nabla_{D^{fe}} l(x_t^{fe}, D^{fe})], \tag{6}$$

where ρ_t is the gradient step at time t, and $\Pi_{\mathbb{D}}$ is the orthogonal projector on the set $\mathbb{D}$. The algorithm converges to a stationary point for a decreasing sequence of ρ_t [3, 9]. Note that the stochastic gradient descent converges but is not guaranteed to converge to a global minimum due to the nonconvexity of the optimization problem [10, 33]. However, experience shows that such a stationary point is sufficiently good for practical applications [13, 32].

To implement our approach, the discrete wavelet transform (DWT) produces set $W_{\text{coe}} = \left\{ lf_i^{fe}, hf_i^{fe} \right\}$ as one low-frequency lf and one high-frequency hf band for each view, and i shows the level of the applied wavelet. A sample of the decomposition step is shown in Fig. 3.

Th joint sparse coding is computed using Eq. (3) for W_{coe} as follows:

$$\min_{A_l\left[\alpha_i^{lf^1}, \dots, \alpha_i^{lf^{FE}}\right]} \frac{1}{2} \sum_{z_f=lf_i^1}^{z_f=lf_i^{FE}} \left\| z_f - D^{z_f}\alpha^{z_f} \right\|_{\iota_2}^2$$
$$+ \lambda_1 \|A_l\|_{\iota_{12}} + \frac{\lambda_2}{2} \|A_l\|_F^2, \tag{7}$$

$$\min_{A_h\left[\alpha_i^{hf^1}, \dots, \alpha_i^{hf^{FE}}\right]} \frac{1}{2} \sum_{z_f=hf_i^1}^{z_f=hf_i^{FE}} \left\| z_f - D^{z_f}\alpha^{z_f} \right\|_{\iota_2}^2$$
$$+ \lambda_1 \|A_h\|_{\iota_{12}} + \frac{\lambda_2}{2} \|A_h\|_F^2, \tag{8}$$

Finally, the corresponding inverse W_{coe} over lf_i^{fe} and hf_i^{fe} based on A_l and A_h is applied to reconstruct the classifiers inputs. Key steps in the feature extraction and fusion steps based on wavelet transform are listed in Table 1.

Note that instead of Steps 2 and 3, we can feed all frequencies into the joint sparse representation simultaneously. The impact of the separation is explored in the experimental results section.

3.3 Classification

For the decision step, the scores of the modal-based classifiers can be combined. The formulation used simultaneously trains the multimodal dictionaries and classifiers under the joint sparsity prior. To classify the classes of multiview problems and obtain a fair comparison, we use the classifiers proposed in [8]. The classifier is based on the joint sparsity before enforcing collaborations among the multiple features and obtains the latent sparse codes as the optimized features for multiclass classification. The performance of these classifiers is studied in the next section. To make the final decision of the classifiers, there are several ways, such as adding corresponding scores and majority voting. In this study, the sum of the score for each feature group is used.

4 Experiments

To evaluate the effectiveness of the proposed system, experiments were conducted on three datasets: IXMAS [47], Animal [24], and NUS-Object [11]. The described method was
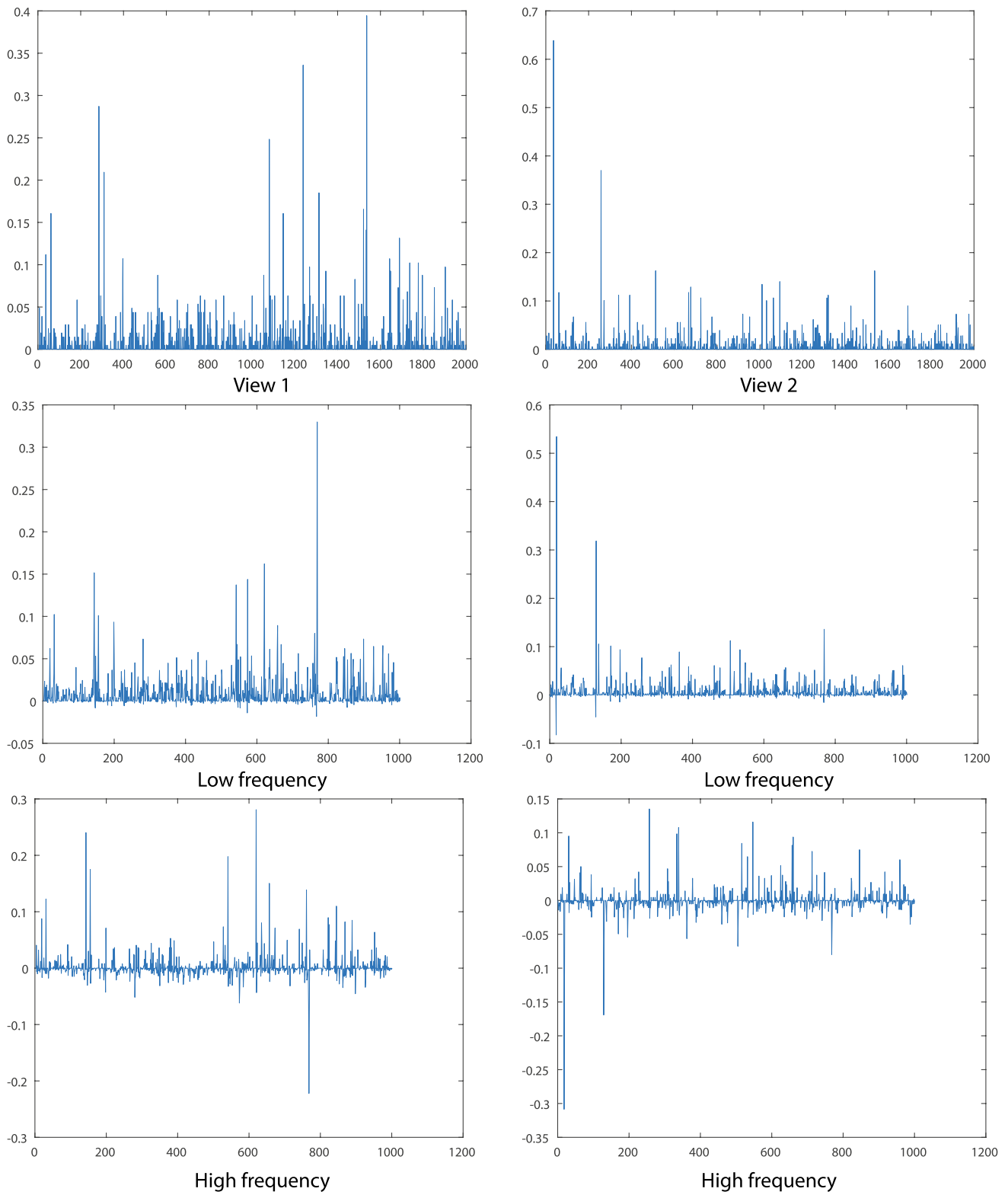
Fig. 3 Sample of decomposition step for two views of IXMAS dataset

compared with state-of-the-art methods including JSRC [41], Wang et al. [45], PLRC [17], AFCDL [18], MDL [8], Grad-KCCA [42], and MvNNcor [48]. The experiments are elaborated in detail in the next subsections.

4.1 Dataset

The performance of proposed method is explored on IXMAS [47], Animal dataset [24], and NUS-Object dataset [11].

IXMAS has images from five different views that can be viewed as a multiview dataset. For each view, there are 11 classes, such as cross arms, scratch head, and check watch. The extracted features and the distribution of the training, validation and test samples are set similarly to [8].

Animal dataset contains 30,475 images with 50 animals classes based on six feature extraction method: color histogram (CH) features, local self-similarity (LSS) features, pyramid HOG (PHOG) features, SIFT features, color SIFT (RGSIFT) features, and SURF features. This can be considered a multiview dataset. The distribution of training, validation, and testing samples are set similar to [48].

NUS-Object has 30,000 images in 31 classes. Methods of CH, color correlogram (CORR), edge direction histogram (EDH), wavelet texture (WT), and block-wise color moments (CM) are applied to extract features. Distribution of training, validation, and testing samples are set similar to [48].

4.2 Experimental setting

The proposed approach was simulated using MATLAB R2019a. All experiments were run on a 64-bit operating system with a CPU E5-2690 and 64.0 GB of RAM. To obtain a fair comparison, we considered all parameters fixed for the fusion method that is introduced in [8] and tested on the databases. In [8], all parameters were carefully analysed. In the joint sparse representation, regularization parameters λ_1 were selected using cross validation in the sets $\{0.01 + 0.005t \mid t \in \{-3, 3\}\}$. The parameter λ_2 was set to zero in most of the experiments as proposed in [8]. Due to its performance in anomaly detection, the Daubechies-2 (db2) wavelet was used to decompose the datasets into a series of

Table 2 Comparison of average accuracies (%) between different fusion methods and our fusion approach on the IXMAS dataset (best value highlighted in bold)

Methods	Average accuracies	Methods	Average accuracies
JSRC [41]	93.60	PLRC [17]	96.20
Wang et al. [45]	93.60	AFCDL [18]	**99.60**
MDL [8]	94.80	Ours	98.20

subbands [23]. To determine the wavelet decomposition levels, three levels of decomposition were performed, as the use of more levels does not affect the detection rates. We also analysed the effects of the levels in Sect. 4.4.

4.3 Results

The proposed method is compared with the other fusion approaches that have been applied to the three datasets. The performance evaluation results (average accuracies) on IXMAS and Animal and NUS-Object are summarized in Tables 2 and 3, respectively. For the IXMAS dataset, we compare our approach with the joint sparse representation classifier (JSRC) [41], Wang et al. [45], Multimodal Task-driven Dictionary Learning (MDL) [8], Pairwise Linear Regression Classification (PLRC) [17], and adaptive fusion and category level dictionary learning (AFCDL) [18]. As shown in Table 2, our approach obtains the second rank in terms of accuracy measurement. Note that our setup for selecting features and training and testing samples is based on [8]. Therefore, it does not lead us to have a fair comparison with other approaches.

Table 3 Comparison of average accuracies (%) between the different fusion methods and our fusion approach on the Animal and NUS-Object datasets (best value highlighted in bold)

Methods	Animal	NUS-Object
MDL [8]	42.20	44.89
GradKCCA [42]	33.33	48.15
MvNNcor [48]	47.69	52.05
Ours	**56**.70	**58**.60

Table 1 Key steps in feature extraction and fusion steps based on our approach

Step 1: Each view is decomposed into subbands using the discrete wavelet transform (DWT). DWT produces a set of subbands with a low-frequency band and a high-frequency band

Step 2: Low frequencies of all views are fused by the joint sparse representation method (Eq. (7))

Step 3: High frequencies of all views are fused by the joint sparse representation method (Eq. (8))

Step 4: Inverse step to reconstruct the classifiers input

The multiview approaches, MDL [8], GradKCCA [42], and MvNNcor [48], are compared in terms of the Animal and NUS-Object datasets, which are two challenging datasets.

As shown in Table 3, our approach achieves the best rank in terms of accuracy measurement.

In addition, to compare our fusion approach performance under the joint sparsity method, one of the best state-of-the-art feature-level fusion algorithms is considered, that is, multimodality dictionary learning [8]. As shown in both tables, we improve the results of the method significantly. Additionally, to analyse the feature space learned, we use the t-SNE algorithm [43] in terms of the IXAMS dataset to project the samples to 2 dimensions, as shown in Fig. 4, our approach distinguishes better than [8].

Finally, the typical computation time for our approach, including wavelet subband extraction and solving Eq. (3) for a given multimodal test sample compared to [8], is shown in Fig. 5 for different dictionary sizes. It shows that as the size increases, the computation time increases linearly for both approaches. The computational cost of our wavelet-based approach is very close to that of [8], where only the wavelet decomposition time is added.

4.4 Impact of decomposition levels

To investigate the effect of the wavelet transform on the classification rates, we perform a series of experiments on the datasets by varying the decomposition levels from 1 to 4. The classification rates are computed and illustrated in Table 4.

It is clarified that after the third level, we cannot see any improvement in the results. We have the best results when we decompose features into three levels.

4.5 Impact of steps 2 and 3

The final experiments are dedicated to exploring our fusion step. To study the impact, we consider classifiers inputs in four states that can occur during fusion. To conduct the first state, we fed only low frequencies into classifiers after their fusion. This process was repeated for the second state with high frequencies. For the third state, we fused both low and high-frequency sets simultaneously without an inverse step. Finally, these were compared with our approach, as shown in Table 5. The results show that our approach obtains the best rank. Additionally, high frequencies do not obtain good results when they are only fed into classifiers. However, they can improve the results significantly when we use them in parallel.

5 Conclusion

This paper proposed a novel fusion approach based on joint sparse representation using the wavelet transform by separating views into low- and high-frequency sets. We
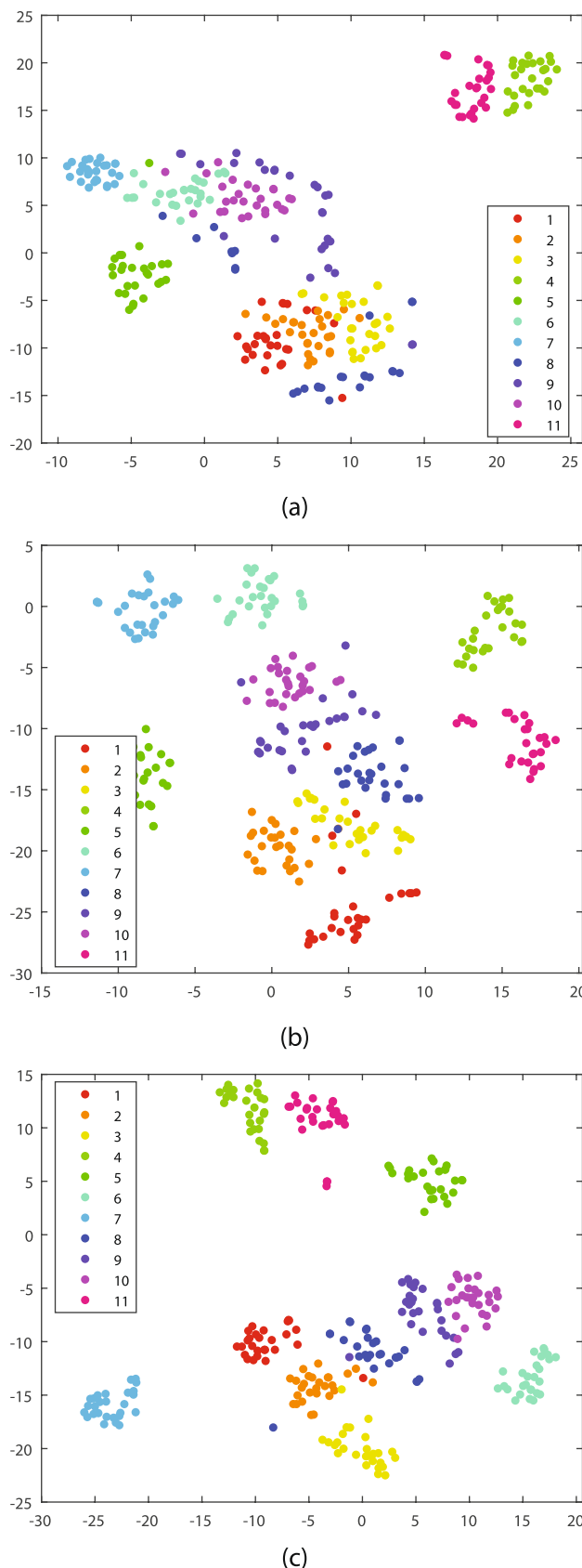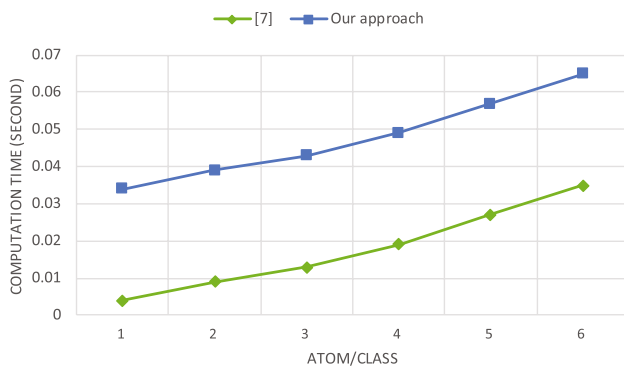

Fig. 4 Visualizations of **a** original data, **b** MDL [8], and **c** our approach using t-SNE on IXAMS dataset

Fig. 5 Computational time for wavelet decomposition and solving optimization problem of Eq. (3) for a given test sample compared with [8]

Table 4 Comparison of impact of wavelet transform levels (best values highlighted in bold)

Levels	IXAMS	Animal	NUS-Object
Level 1	97.50	52.60	53.70
Level 2	97.95	55.401	57.20
Level 3	**98.20**	**56.70**	**58.60**
Level 4	96.90	50.30	51.20

Table 5 Comparison of impact of our fusion approach (best values highlighted in bold)

Fusion approach	IXAMS	Animal	NUS-Object
Only lf_i^{fe}	96.80	45.50	49.40
Only hf_i^{fe}	65.50	30.50	39.20
Both simultaneously	97.60	53.30	55.50
Ours	**98.20**	**56.70**	**58.60**

considered three datasets that included data with multiple views. In the first step, wavelet transform was applied to the different views, and approximation and detail coefficients were obtained. Then, we considered a fusion step at the feature level using the joint sparse representation tool. Low and high frequencies were fed into the fusion method separately. Using an inverse discrete wavelet transform, we reconstructed a new space based on both the low and high frequencies after applying the fusion method. To make a decision, the output of the step was fed into classifiers. The presented approach was tested on the three datasets. The results produced by this method were generally better in terms of the datasets than the state-of-the-art results.

Based on our experiments, we can claim that separating features can improve the results of fusion methods. Therefore, as future work, we aim to customize our approach based on other fusion approaches. Additionally, we aim to develop the model for a deep approach when the size of the feature vectors is sufficient for the purpose. Other filter banks can be applied for comparison with wavelet transforms.

References

1. Abavisani M, Patel VM (2018) Multimodal sparse and low-rank subspace clustering. Inf Fusion 39:168–177
2. Adam K, Al-Maadeed S, Akbari Y (2022) Hierarchical fusion using subsets of multi-features for historical arabic manuscript dating. J Imaging 8(3):60
3. Aharon M, Elad M (2008) Sparse and redundant modeling of image content using an image-signature-dictionary. SIAM J Imaging Sci 1(3):228–247
4. Akbari Y, Nouri K, Sadri J et al (2017) Wavelet-based gender detection on off-line handwritten documents using probabilistic finite state automata. Image Vis Comput 59:17–30
5. Akbari Y, Al-Maadeed S, Adam K (2020) Binarization of degraded document images using convolutional neural networks and wavelet-based multichannel images. IEEE Access 8:153,517-153,534
6. Akbari Y, Hassen H, Subramanian N, et al (2020) A vision-based zebra crossing detection method for people with visual impairments. In: 2020 IEEE international conference on informatics, IoT, and enabling technologies (ICIoT), IEEE, pp 118–123
7. Akbari Y, Almaadeed N, Al-Maadeed S et al (2021) Applications, databases and open computer vision research from drone videos and images: a survey. Artif Intell Rev 54(5):3887–3938
8. Bahrampour S, Nasrabadi NM, Ray A et al (2015) Multimodal task-driven dictionary learning for image classification. IEEE Trans Image Process 25(1):24–38
9. Bottou L (2010) Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT'2010. Springer, pp 177–186
10. Bottou L, Bousquet O (2007) The tradeoffs of large scale learning. Adv Neural Inf Process Syst 20:351–368
11. Chua TS, Tang J, Hong R, et al (2009) Nus-wide: a real-world web image database from national university of singapore. In:

Proceedings of the ACM international conference on image and video retrieval, pp 1–9

12. Cotter SF, Rao BD, Engan K et al (2005) Sparse solutions to linear inverse problems with multiple measurement vectors. IEEE Trans Signal Process 53(7):2477–2488

13. Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process 15(12):3736–3745

14. Elharrouss O, Akbari Y, Almaadeed N, Al-Maadeed S (2022) Backbones-review: Feature extraction networks for deep learning and deep reinforcement learning approaches. arXiv preprint arXiv:2206.08016

15. Elharrouss O, Almaadeed N, Al-Maadeed S et al (2020) Image inpainting: a review. Neural Process Lett 51(2):2007–2028

16. Feng CM, Xu Y, Li Z, et al (2019) Robust classification with sparse representation fusion on diverse data subsets. arXiv:1906.11885

17. Feng Q, Zhou Y, Lan R (2016) Pairwise linear regression classification for image set retrieval. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4865–4872

18. Gao Z, Xuan HZ, Zhang H et al (2019) Adaptive fusion and category-level dictionary learning model for multiview human action recognition. IEEE Internet Things J 6(6):9280–9293

19. Gui J, Tao D, Sun Z et al (2014) Group sparse multiview patch alignment framework with view consistency for image classification. IEEE Trans image Process 23(7):3126–3137

20. Hall DL, Llinas J (1997) An introduction to multisensor data fusion. Proc IEEE 85(1):6–23

21. Hu S, Yan X, Ye Y (2020) Joint specific and correlated information exploration for multi-view action clustering. Inf Sci 524:148–164

22. Kan M, Shan S, Zhang H et al (2015) Multi-view discriminant analysis. IEEE Trans Pattern Anal Mach Intell 38(1):188–194

23. Kanarachos S, Christopoulos SRG, Chroneos A et al (2017) Detecting anomalies in time series data via a deep learning algorithm combining wavelets, neural networks and hilbert transform. Expert Syst Appl 85:292–304

24. Lampert CH, Nickisch H, Harmeling S (2009) Learning to detect unseen object classes by between-class attribute transfer. In: 2009 IEEE conference on computer vision and pattern recognition, IEEE, pp 951–958

25. Lee DD, Seung HS (1999) Learning the parts of objects by non-negative matrix factorization. Nature 401(6755):788

26. Li B, Yuan C, Xiong W et al (2017) Multi-view multi-instance learning based on joint sparse representation and multi-view dictionary learning. IEEE Trans Pattern Anal Mach Intell 39(12):2554–2560

27. Li J, Zhang B, Zhang D (2017) Joint discriminative and collaborative representation for fatty liver disease diagnosis. Expert Syst Appl 89:31–40

28. Li J, Zhang D, Li Y et al (2017) Joint similar and specific learning for diabetes mellitus and impaired glucose regulation detection. Inf Sci 384:191–204

29. Li J, Zhang B, Lu G et al (2019) Generative multi-view and multi-feature learning for classification. Inf Fusion 45:215–226

30. Li SY, Jiang Y, Zhou ZH (2014) Partial multi-view clustering. In: Twenty-Eighth AAAI conference on artificial intelligence

31. Liu H, Liu L, Le TD et al (2017) Nonparametric sparse matrix decomposition for cross-view dimensionality reduction. IEEE Trans Multimed 19(8):1848–1859

32. Mairal J, Elad M, Sapiro G (2007) Sparse representation for color image restoration. IEEE Trans Image Process 17(1):53–69

33. Mairal J, Bach F, Ponce J, et al (2010) Online learning for matrix factorization and sparse coding. J Mach Learn Res 11(1):19–60

34. Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 11(7):674–693

35. Parikh N, Boyd S, et al (2014) Proximal algorithms. Found Trends® Optim 1(3):127–239

36. Rakotomamonjy A (2011) Surveying and comparing simultaneous sparse approximation (or group-lasso) algorithms. Signal Process 91(7):1505–1526

37. Ross AA, Govindarajan R (2005) Feature level fusion of hand and face biometrics. In: Biometric technology for human identification II. International Society for Optics and Photonics, pp 196–204

38. Ruta D, Gabrys B (2000) An overview of classifier fusion methods. Comput Inf Syst 7(1):1–10

39. Shao L, Liu L, Yu M (2016) Kernelized multiview projection for robust action recognition. Int J Comput Vis 118(2):115–129

40. Shariatmadari S, Emadi S, Akbari Y (2020) Nonlinear dynamics tools for offline signature verification using one-class gaussian process. Int J Pattern Recognit Artif Intell 34(01):2053,001

41. Shekhar S, Patel VM, Nasrabadi NM et al (2013) Joint sparse representation for robust multimodal biometrics recognition. IEEE Trans Pattern Anal Mach Intell 36(1):113–126

42. Uurtio V, Bhadra S, Rousu J (2019) Large-scale sparse kernel canonical correlation analysis. In: International conference on machine learning, PMLR, pp 6383–6391

43. Van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605

44. Varshney PK (1997) Multisensor data fusion. Electron Commun Eng J 9(6):245–253

45. Wang H, Kläser A, Schmid C et al (2013) Dense trajectories and motion boundary descriptors for action recognition. Int J Comput Vis 103(1):60–79

46. Wang W, Arora R, Livescu K, et al (2015) On deep multi-view representation learning. In: International conference on machine learning, pp 1083–1092

47. Weinland D, Boyer E, Ronfard R (2007) Action recognition from arbitrary views using 3d exemplars. In: 2007 IEEE 11th international conference on computer vision, IEEE, pp 1–7

48. Xu J, Li W, Liu X, et al (2020) Deep embedded complementary and interactive information for multi-view classification. In: Proceedings of the AAAI conference on artificial intelligence, pp 6494–6501

49. Yang M, Zhang L, Zhang D, et al (2012) Relaxed collaborative representation for pattern classification. In: 2012 IEEE conference on computer vision and pattern recognition, IEEE, pp 2224–2231

50. Yuan XT, Liu X, Yan S (2012) Visual classification with multitask joint sparse representation. IEEE Trans Image Process 21(10):4349–4360

51. Zhang H, Zhang Y, Nasrabadi NM et al (2012) Joint-structured-sparsity-based classification for multiple-measurement transient acoustic signals. IEEE Trans Syst Man Cybern Part B (Cybern) 42(6):1586–1598

52. Zhao Z, Lu H, Deng C, et al (2016) Partial multi-modal sparse coding via adaptive similarity structure regularization. In: Proceedings of the 24th ACM international conference on Multimedia, ACM, pp 152–156