International Journal of Web-Based Learning and Teaching Technologies
Volume 18 • Issue 2

Analysis of Multimedia Feature Extraction Technology in College Vocal Performance Teaching Mode Based on Multimodal Multimedia Information

Weijuan Nie, Zhengzhou Preschool Education College, China*
Wan Ng, University of Technology Sydney, Australia

ABSTRACT

This article is based on the application research of multimedia feature extraction technology in the development of vocal performance teaching in universities. Combining with the core image and sound modules in feature extraction technology, this article proposes an application model of multimedia feature extraction technology based on image HOG algorithm and Mel spectrum for vocal audio recognition in vocal performance teaching in universities. Experiments have shown that the features extracted by this method can not only effectively identify the styles of different types of singing works, but also recognize the personality characteristics of singers. At the same time, it can effectively reduce the misclassification rate caused by noise interference, thereby improving the recognition rate.

KEYWORDS

Aesthetic Education, LBP Algorithm, Mel Spectrum, Multimedia Feature Extraction Technology, Vocal Performance Teaching

INTRODUCTION

Over time, changes have occurred within science, technology, and the aesthetic level of the public . An evolution is also taking place in the requirements surrounding aesthetic education. In recent years, music reality shows like The Voice of China and Mask Singer boast the highest ratings in their broadcast period. These programs are closely related to music, demonstrating the broad prospects and significance of the vocal performance major in colleges and universities (Hu, 2022).

Regarding vocal performance teaching in colleges and universities, it is difficult to achieve the desired teaching quality and effect solely through teacher-student classroom interaction and students' autonomous after-class learning time (Jin-ping et al., 2012). The teaching modes related to vocal performance have witnessed significant changes with the integration of digital multimedia technologies

DOI: 10.4018/IJWLTT.329604 *Corresponding Author

(Liu, 2022). Deep integration, for example, applies various forms of multimedia technology to vocal music, including music analysis, music education, music score following, sound mixing, vocal performance robot, deep learning, and picture and video soundtrack.

The practical application of technologies has enhanced the development of vocal music education in colleges and universities (Liu, 2021). For instance, multimodal teaching integrates multiple symbolic modes (e.g., images, video, and language text) into teaching activities. These tools can encourage students to participate in the learning process. The popularization and improvement of multimedia teaching equipment in colleges and universities provides convenient conditions for the introduction of mobile multimodal teaching into vocal music teaching in colleges and universities.

Much of the existing research focuses on how to apply multimedia technology to the vocal music teaching mode in colleges and universities. Few studies analyze the multimedia feature extraction technology of vocal music teaching mode in colleges and universities based on multimodal multimedia information. Therefore, this article proposes an application mode of multimedia feature extraction technology based on image HOG algorithm and Mel spectrum in vocal music performance teaching in colleges and universities.

LITERATURE REVIEW

Status Quo: Vocal Performance Teaching Mode in Colleges and Universities in China

With rapid developments in science and technologies, more products have computer technologies at their core. These, in turn, play a leading role in our daily lives. Multimedia technology became one of the strongest developments in the field of information technology (Atrey et al., 2010). With the help of multimedia technology, users become creators. That is, one person can capture a short video using one mobile phone or computer. Thus, the relevant short video teaching industry chain also appears within reach (Bai, 2022). Behind any excellent short video is the shadow of college vocal performance education (Crawford, 2017). Per Table 1, creators of many high-quality short videos often utilize a professional team (e.g., copywriting, vocals, music, and performance and value of videos).

The reform and development of vocal music teaching in colleges and universities cannot be separated from the innovation of teaching methods. At the same time, it also needs to introduce diversified teaching methods to assist in the development of teaching activities. Multimedia technology is one resource for improving vocal music teaching.

Table 1. Vibrato short video powder fastest TOP10

Ranking	Short Video Account Name	Powder Increase (10,000)
1	People's Daily	6,615
2	CCTV News	6,422
3	Poisonous Tongue Film	4,673
4	Sichuan Observation	3,647
5	Zheng Jianpeng, the Big Wolf Dog & Yanzhen Couple	2,944
6	Crazy Sisters	2,899
7	Crazy Xiaoyangge	2,726
8	I'm Granny Tian	2,610
9	Monkey Said the Car	2,348
10	Cloth Exploration	2,125

The most recent reform of vocal music education was carried out in 2017, in which the Chinese Ministry of Education issued a guiding opinion that "aesthetic education such as vocal music and performance is not only subject education in colleges and universities, but also people-oriented education" (Heittola et al., 2013). Many achievements have been made in vocal music aesthetic education in colleges and universities in China since the reform and innovation of aesthetic education. For example, teachers are continuing their education, new and personalized teaching modes are utilized, teaching concepts are optimized, and teaching methods are diversified (Fan, 2021). Still, there are many problems to overcome in vocal music education in colleges and universities.

The first problem involves vocal performance courses in colleges and universities. The core strength of vocal performance education in colleges and universities has served as a foundation for the implementation of vocal performance teaching in colleges and universities (Maddumala, 2020). Unfortunately, many vocal performance courses are reduced to electives due to the inaccurate positioning of students' talent training plan in comprehensive universities (Mansoorizadeh & Charkari, 2010).

The second problem centers around teachers, who play a vital role in the effective development of vocal performance education in colleges and universities (Nauman et al., 2020). As shown in Table 2, the faculty of vocal performance in colleges and universities is relatively weak. Quantitative analysis alone cannot meet the increasing teaching demand of colleges and universities or its association with teaching quality (Pandit et al., 2019).

The third is the problem of teaching mode. Multimedia networks are gaining popularity due to rapid developments in technology. This, in turn, presents innovative opportunities to vocal performance education in colleges and universities (Li, 2016). The teaching mode of vocal performance in colleges and universities has been optimized under the guidance of "Spring Breeze." However, the root of the teaching mode is still too single. It is still a "one-to-many" + "one-to-one" dual-mode implementation or one-to-one practice skill drills in small classes and one-to-many theoretical knowledge lectures in large classes (Yang & Chong, 2021).

A fourth problem focuses on the teaching activity rating system. At present, the evaluation system of vocal music teaching in colleges and universities in China is not very scientific. The evaluation content is incomplete and too much attention is paid to the investigation of students' vocal music skills (Bonastre et al., 2017). Other factors, such as emotion and attitude outside the skills, are ignored. Thus, we cannot test the overall quality of students.

Characteristics of Vocal Performance Teaching Mode in Foreign Universities

There are three characteristics of high-efficiency vocal performance teaching mode in foreign countries. These include: (1) artistic integrity; (2) coordination of singing; and (3) emphasis on voice training (Zong & Huang, 2021).

Table 2. Number of aesthetic education teachers and enrollment in colleges and universities

Year	Number of Art Course Teachers (10,000)	Growth Rate	Undergraduate Enrollment (10,000)	Growth Rate
2015	59.9	----	366.01	----
2016	63.4	5.84	374.11	2.18
2017	68.7	8.36	372.01	-0.53
2018	71.7	4.37	422.16	13.48
2019	74.8	4.32	431.29	2.16
2020	77.8	4.01	443.10	2.74
2021	83.0	6.68	444.60	0.34

The wholeness of art, which is a highly concentrated description of the integration and unity of natural environment, gives people a visual feeling of beauty. This is why the ideological education and spiritual edification we receive within quality education are collectively referred to as "aesthetic education" (Zhang & Wang, 2022). Its essence is found in the process that human beings reverse explore, pursuing the law and essence of beauty through natural harmony. In the teaching of artistic skills, vocal music art pays attention to the cultivation of college students' personal qualities, humanistic atmosphere, and comprehensive understanding (Zhao & Liu, 2022).

The essence of coordination in singing is to make $1 + 1 > 2$. There are many ways to achieve coordination in vocal music, including perfect coordination between piano art and vocal music (which we are familiar with) and the singers (Jay Chou and Ronghao Li), who are trendy musicians and pianists. It is one of the most common artistic means in piano vocal performance. Although each has different artistic expression techniques, the appeal of piano performance can be radiated to a greater extent by setting off the melody of vocal performance to more clearly express the singer's emotion, the creative purpose of the work, and the artistic value of the performance and performers (Bokiev et al., 2018).

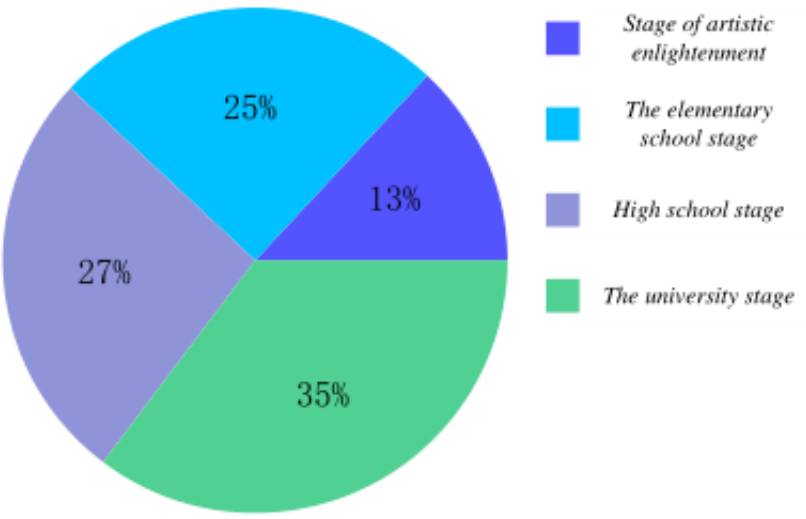
Regarding voice training and vocal performance, it is important to "strike while the iron is hot." Vocal performers attach great value to the practice of individual vocalization and improvement of vocalization skills. Foreign vocal performance teaching has gradually increased the promotion of investment in aesthetic education courses at various educational stages (see Figure 1). In terms of skills, it pays more attention to students' layered practice of high-pitched and low-pitched voices, which makes students' vocalization more three-dimensional and layered. This teaching mode also enables students to establish a strong vocal awareness and skill accomplishment from the learning era of vocal performance, setting a solid base for subsequent students' artistic development.

METHODOLOGY

Histogram of Gradient and Feature Extraction Technology

Scholars aim to study methods to extract features and design a visual model system that can be processed and recognized by computers. In general, a corresponding visual model system is constructed

Figure 1. Proportion of investment in various stages of foreign aesthetic education courses



by using different feature layers to represent patterns that can be recognized by computers (Powell et al., 2020). The first problem to be considered in the visual model system is the image scale problem or visual multi-scale analysis problem (Rakotomamonjy & Gasso, 2015). The idea is that when we observe an object with our eyes and the distance between the object and observer is changing, the pixel size of the same object converts into an image change. We must identify a way to find the essential characteristics of the key object in the image with different pixel sizes. This visual analysis method is called visual multi-scale analysis.

The feature extraction methods of basic acoustics can be divided into the following categories:

1. **Recognition Based on Low-Order Features in Time-Frequency Domain:** Eronen and Malkin (2013) recognized low-order features through the simple calculation of time-domain signals or frequency-domain signals after Fourier transform.
2. **Spectrum Characteristics:** Spectrum characteristics refer to the corresponding amplitude spectrum or energy spectrum of the signal as calculated through Fourier transform and a group of filters. It can reflect the dynamic spectrum characteristics of audio.
3. **Cepstrum Feature:** This feature is used to roughly capture the spectrum envelope. MFCC is the most common cepstrum feature in audio scene recognition.
4. **Phonation Feature:** The phonation feature assumes that the signal contains harmonic components. In the field of music retrieval, voice features have been widely used. In recent years, it has also received some attention in the field of audio scene recognition and audio event detection. For example, Geiger generates a set of features to measure the fundamental frequency characteristics (Mera et al., 2014).

Multimedia Feature Extraction Technology

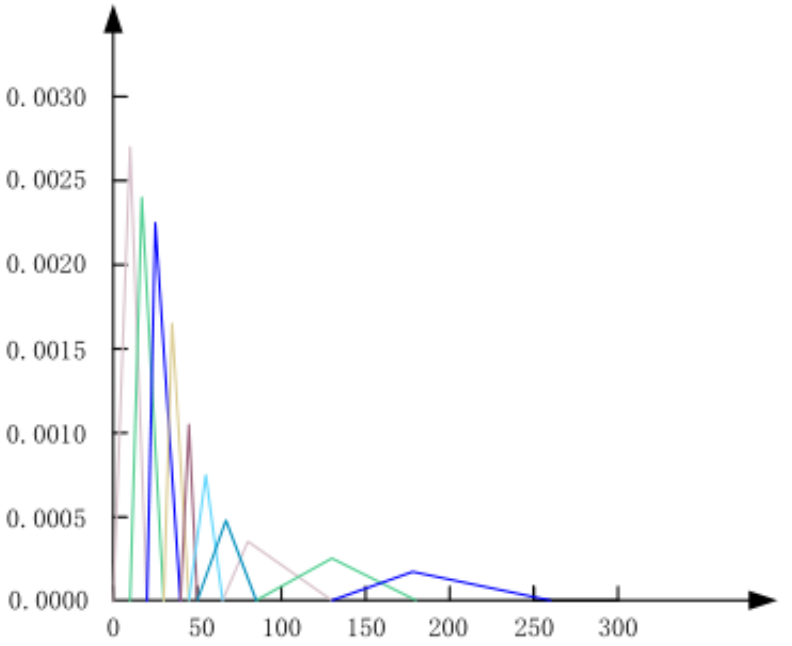
The innovation of vocal performance teaching mode in colleges and universities also benefits from the leading-edge research and development of related technologies. This, in turn, supported the rise of feature extraction technologies or the special diagnosis extraction of samples from the initial measurement data of samples through machine autonomous learning and recognition of mathematical models and sound processing. The extracted information is output separately to establish accurate, non-redundant derived values and promote the subsequent use and learning of sample information (Fu, 2020).

In this article, the local binary patterns algorithm (LBP) and the Mel spectrum and Mel cepstrum techniques are used as explanation targets.

The LBP algorithm is often used to extract the texture features of visual images. It has the characteristics of gray invariance. This feature does not allow the intensity of illumination to affect the image area. In addition, it will not cause the pixel value of a pixel area to expand in equal ratio because the strong light hits the pixel area (Chen et al., 2018). The algorithm re-encodes the binary by using the ratio of the central pixel to the domain pixel in a certain pixel area. The final binary size relationship will not be affected by other factors. The overall process is shown in Figure 2. The "local" is viewed as the 3*3 sliding area of the algorithm calculation unit. Then, the central pixel value of the calculation unit is taken as the threshold value of the eight adjacent fields. Binarizing the pixels sets the gray value of the pixels on the image to 0 or 255. Thus, the visual effect of the 3*3 unit is black and white.

Mel spectrum is an indispensable part of vocal music performance. Atlas is usually transformed into Mel spectrum by Mel-scale filter banks in multimedia feature extraction technology. Then, the feature is processed. The pitch heard through our ears is linearly related to the logarithm of the fundamental frequency of the sound. The principle of the Mel scale filter bank is that when the difference of Mel frequency between two audio segments of the detected object AB is N times, the tone perceived by the human ear is about N times. The Mel rising curve changes slowly when the Mel

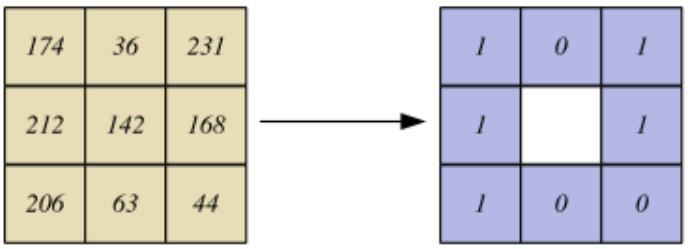
Figure 2. Local illustration of LBP algorithm



frequency is high. The slope of the curve is less than 1 (vice versa, it is greater than 1). Thus, it is proved that human ears will be more sensitive to low-frequency vocal music, while high-frequency music is relatively slow. Mel cepstrum is a spectrum obtained by cepstrum analysis based on Mel spectrum or taking the logarithm of data results and performing discrete cosine transform (see Figure 3).

Short-time Fourier transform, which is suitable for spectrum analysis of slow time-varying signals, has been used in audio and image analysis and processing. The method divides the signal into frames. Then, it carries out the Fourier transform on each frame. The speech signal is stable in a short time; thus, it can be divided into frames. The Fourier transform of a certain frame can be calculated, which is the short-time Fourier transform.

Figure 3. Mel scale filter



$$(10110011)_{10}=179$$

Optimization: Vocal Performance Teaching Mode in Chinese Universities

The teaching evaluation system in China can be optimized. Under the current teacher-centered evaluation system, teachers must adopt objective and scientific evaluation standards, reduce the influence of a teacher's individual factors on student evaluation, and reflect daily teaching. In addition, the forms of evaluation should be diversified; thus, students' basic knowledge and skills of vocal music cannot be examined purely and the learning process, feelings, attitudes, and values should be evaluated to understand students' comprehensive quality. By analyzing the present situation of vocal performance teaching in colleges and universities in China, as well as the teaching characteristics of foreign countries, this article puts forward optimization schemes for the teaching mode of vocal performance in China in combination with an actual situation.

Optimize Vocal Performance Teaching Curriculum

It is beneficial to learn from authoritative professional colleges. When learning from them, we can launch a survey questionnaire on students' artistic perception in school. Then, we can draw up a personalized and customized aesthetic education course in line with colleges and universities. In the acceptance of teaching achievements, various modes are adopted. Theoretical knowledge, vocal performance skills, and emergency response ability should be checked. Vocal music teaching is both technical and artistic. If students want to learn vocal music, they need tacit cooperation with teachers. Before teaching, teachers should master each student's voice characteristics and adopt teaching methods suitable for student development according to their respective voice characteristics. Teachers should make a detailed development plan for students, communicating with students so they can define their individual development direction.

Strengthen the Construction of Teaching Staff in Multiple Dimensions

First, as the reserve force of teachers in the aesthetic education system, music major colleges and normal colleges can support the aesthetic education talents program. This includes a focus on professional training and teaching skills, providing a large number of qualified talents for the aesthetic education in primary and secondary schools in China. Second, to increase the talent introduction policy in colleges and universities, we should both introduce cutting-edge vocal performance theory talents with high academic qualifications and open the doors of colleges and universities with experienced vocal performance skills. Foreign art colleges are targeted. Thus, most of the high-quality art design colleges are independent and specialized, ensuring the professionalism of teaching. Additionally, foreign art learning is open-minded, stressing concepts and encouraging students to think differently. Students are encouraged to study abroad, promoting the benefits of a social atmosphere and artistic backgrounds. According to their teaching situation, universities conduct daily teaching seminars, analyze teaching results regularly, and combine daily teaching with research to enhance vocal performance teaching under the guidance of science and specialty.

Reform the Teaching Mode

This module is the most difficult to adopt. The continuous expansion of enrollment in colleges and universities leads to a growing number of teachers and students. In this case, we should implement the core mode of teaching "teaching in accordance with student aptitude." Most students will be graded according to their vocal performance cognition, level, degree of interest, and personality. The teaching work will be carried out hierarchically and pertinently.

The education should make full use of the current multimedia network technologies (e.g., cloud teaching, cloud acceptance, and cloud practice) to equip students with artistic creation and vocal performances. Vocal music teaching in colleges and universities can be carried out through a regular exchange between teachers and students. That is, students learn and prepare lessons for the sections they excel at or need support.

Everyone in the classroom should brainstorm about this problem after explaining and analyzing this module. This is conducive to the mastery of knowledge between teachers and students. It is also convenient for teachers to know students' learning situations.

RESULTS AND DISCUSSION

Analysis: The Application of Multimedia Feature Extraction Technology to the Teaching Mode of Vocal Music Performance

There are several reasons why multimedia feature extraction technology applies to the teaching mode of vocal music performance in colleges and universities.

First, the analysis of college vocal performance teaching mode is inseparable from the audio signal. For instance, sound can convey too much information (e.g., news, feelings, and artistic conception). The information is, essentially, another form of data. It is important to accurately extract and analyze the audio features to analyze the college vocal performance teaching mode. The current reinforcement learning algorithm can play a strong auxiliary role in the extraction of audio features. This greatly improves the scientific results of collection and reinforcement learning output, ensuring that the extracted feature data is accurate, effective, and can be used by vocal performance teaching in colleges and universities (Zhao & Liu, 2022).

Second, the audio in vocal performance teaching in colleges and universities also has professional characteristics, such as loudness, which is a psychological measure of a person's voice supervisor. In general, when the sound frequency is certain, the higher the sound intensity and the greater the loudness. The relationship between loudness, frequency, and sound intensity needs to be judged from the equal loudness curve. These professional feature outputs can be solved by the existing multimedia feature extraction technology.

Third, multimedia technology has been integrated into vocal performance teaching in colleges and universities. The teaching mode in colleges and universities has become more intelligent and scientific; thus, the digital construction of the evaluation system for vocal performance teaching will change. These changes should be assisted by multimedia feature extraction technologies to enforce strong information reliability.

Application Model: Multimedia Feature Extraction Technology and the Analysis of Vocal Performance Teaching Mode

The model establishment of this college vocal performance teaching mode analysis is mainly based on the performance vision and vocal audio identification.

The HOG feature extraction algorithm is an effective approach to analyzing visual performance. The term "HOG direction gradient histogram" is easy to understand. By detecting the density of edges or gradients in different azimuths, the HOG algorithm can accurately capture the essence of an image and its shape. Besides the same advantage of gray scale invariance, it has geometric invariance. The HOG feature is formed by calculating and counting the gradient direction histogram of the local area of the image (like the idea of LBP). The HOG feature extraction divides the image into several whole parts (see Figure 4), in which the largest Window level is composed of an integer number of blocks. The blocks are composed of an integer number of units (Rakotomamonjy & Gasso, 2015).

On the other hand, the recognition of vocal music audio is based on the Mel spectrum. This is a result of the point multiplication of the spectrum and several Mel filters. The Mel scale, a nonlinear transformation metric of Hertz (Hz), can effectively distinguish between similar frequencies like as 6500 Hz and 6800 Hz. Using the Mel scale, the Mel spectrum can precisely represent the acoustic features of vocal music, identifying the unique characteristics of each performance. Multiplying these dots is shown in Figure 5. This obtains the Mel spectrum.

Figure 4. HOG structure display diagram

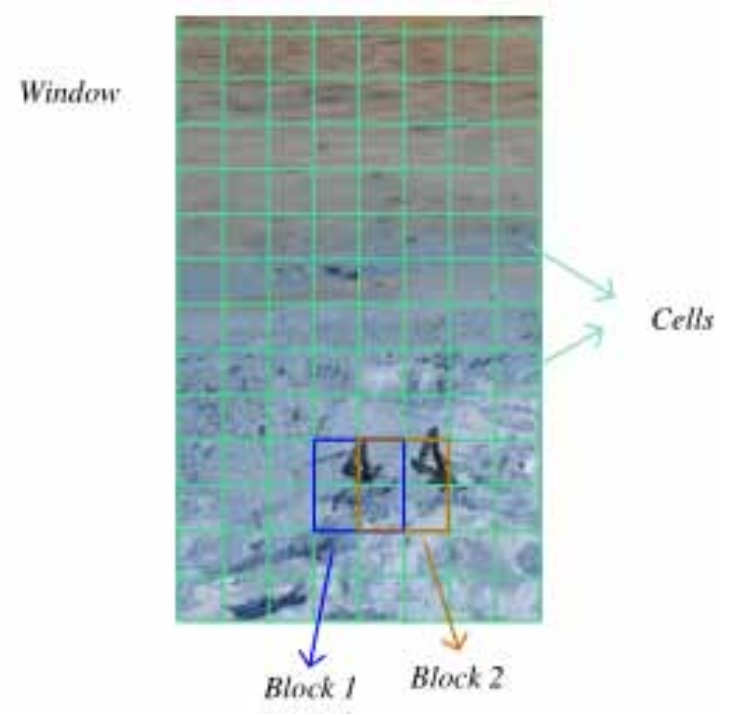


Figure 5. Code fragment of dot multiplication of mel filter bank

```
bin = numpy.floor((NFFT + 1) * hz_points / sample_rate)

fbank = numpy.zeros((nfilt, int(numpy.floor(NFFT / 2 + 1))))

for m in range(1, nfilt + 1):

    f_m_minus = int(bin[m - 1])

    f_m = int(bin[m])

    f_m_plus = int(bin[m + 1])

    for j in range(f_m_minus, f_m):

        fbank[m - 1, j] = (j - bin[m - 1]) / (bin[m] - bin[m - 1])

    for j in range(f_m, f_m_plus):

        fbank[m - 1, j] = (bin[m + 1] - j) / (bin[m + 1] - bin[m])
```

At present, the efficient vocal performance teaching mode is based on vocal music and a coordinated performance. To restore the tone, intonation, demeanor, and movements of the performer, video and audio are used for double recording. Then, the teacher makes a split analysis on the performer. This kind of analysis tests the teacher's experience and performance of the recording equipment. However, the analysis results are imperfect. Therefore, images and audio are processed by multimedia feature extraction technology, allowing the teachers to coach performers more rationally, comprehensively, and scientifically.

CONCLUSION

This article discusses the application of multimedia feature extraction technology in college vocal music performance teaching. It introduces the development of multimedia feature extraction technology, as well as its core image and sound modules. Then, it analyzes the current situation of vocal music performance teaching in Chinese universities, foreign teaching characteristics, and optimization schemes. Based on these analyses, the article proposes an application mode for multimedia feature extraction technology based on the HOG algorithm for image processing and Mel spectrum for sound processing. The experiment shows that this method can effectively identify the style of different singing works and the singer's personality characteristics. It can also reduce errors caused by noise interference, thus improving the recognition rate. Furthermore, the application of this method in practical classroom teaching can enhance students' learning interest, encourage the correct use of one's voice, and cultivate strong singing habits.

In conclusion, the proposed application mode for multimedia feature extraction technology provides a practical and effective approach to improve the quality of vocal music performance teaching in colleges and universities.

DATA AVAILABILITY

The figures and tables used to support the findings of this study are included in the article.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING STATEMENT

This work was not supported by any funds.

ACKNOWLEDGMENT

The authors would like to thank those who have contributed to this research.

REFERENCES

Atrey, P. K., Hossain, M. A., El Saddik, A., & Kankanhalli, M. S. (2010). Multimodal fusion for multimedia analysis: A survey. *Multimedia Systems*, *16*(6), 345–379. doi:10.1007/s00530-010-0182-0

Bai, J. (2022). Optimized piano music education model based on multimodal information fusion for emotion recognition in multimedia video networks. *Mobile Information Systems*, *2022*, 1–12. doi:10.1155/2022/1882739

Bonastre, C., Muñoz, E., & Timmers, R. (2017). Conceptions about teaching and learning of expressivity in music among higher education teachers and students. *British Journal of Music Education*, *34*(3), 277–290. doi:10.1017/S0265051716000462

Chen, Y., Chen, C., Wu, S., & Lo, C. (2018). A two-step approach for classifying music genre on the strength of AHP weighted musical features. *Mathematics*, *7*(1), 19. doi:10.3390/math7010019

Crawford, R. (2017). Rethinking teaching and learning pedagogy for education in the twenty-first century: Blended learning in music education. *Music Education Research*, *19*(2), 195–213. doi:10.1080/14613808.2016.1202223

Fan, Y. (2021). Application of computer technology in vocal music teaching. *Journal of Physics: Conference Series*, *1881*(2), 022050. doi:10.1088/1742-6596/1881/2/022050

Fu, L. (2020). Research on the reform and innovation of vocal music teaching in colleges. *Region-Educational Research and Reviews*, *2*(4), 37–40. doi:10.32629/rerr.v2i4.202

Heittola, T., Mesaros, A., Eronen, A., & Virtanen, T. (2013). Context-dependent sound event detection. *EURASIP Journal on Audio, Speech, and Music Processing*, *2013*(1), 1–13. doi:10.1186/1687-4722-2013-1

Hu, Y. (2022). Music emotion research based on reinforcement learning and multimodal information. *Journal of Mathematics*, *2022*, 1–9. doi:10.1155/2022/2446399

Jin-ping, Y., Xi-mei, H., & Xiao-yun, X. (2012). *Image data mining technology of multimedia: Future wireless networks and information systems*. Springer. doi:10.1007/978-3-642-27326-1_49

Li, M. (2016). Smart home education and teaching effect of multimedia network teaching platform in piano music education. *International Journal of Smart Home*, *10*(11), 119–132. doi:10.14257/ijsh.2016.10.11.11

Liu, S. (2021). Multimedia interactive system of vocal music teaching based on voice recognition. *13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)* (pp. 599–602). IEEE. doi:10.1109/ICMTMA52658.2021.00138

Liu, T. (2022). Multidimensional nonlinear landscape planning based on parameter feature extraction and multimedia technology. *Advances in Multimedia*, *2022*, 1–7. doi:10.1155/2022/5477516

Maddumala, V. R. (2020). A weight based feature extraction model on multifaceted multimedia big data using convolutional neural network. *Ingénierie des Systèmes d'Information, 25*(6).

Mansoorizadeh, M., & Charkari, N. M. (2010). Multimodal information fusion application to human emotion recognition from face and speech. *Multimedia Tools and Applications*, *49*(2), 277–297. doi:10.1007/s11042-009-0344-2

Mera, D., Batko, M., & Zezula, P. (2014). Towards fast multimedia feature extraction: Hadoop or storm. *2014 IEEE International Symposium on Multimedia* (pp. 106–109). IEEE. doi:10.1109/ISM.2014.60

Nauman, A., Qadri, Y. A., Amjad, M., Zikria, Y. B., Afzal, M. K., & Kim, S. W. (2020). Multimedia Internet of things: A comprehensive survey. *IEEE Access : Practical Innovations, Open Solutions*, *8*, 8202–8250. doi:10.1109/ACCESS.2020.2964280

Pandit, V., Amiriparian, S., & Schmitt, M. (2019). Big data multimedia mining: Feature extraction facing volume, velocity, and variety. *Big Data Analytics for Large-Scale Multimedia Search, 61*.

Powell, B., Hewitt, D., Smith, G. D., Olesko, B., & Davis, V. (2020). Curricular change in collegiate programs: Toward a more inclusive music education. *Visions of Research in Music Education*, *35*(1), 16.

Rakotomamonjy, A., & Gasso, G. (2015). Histogram of gradients of time-frequency representations for audio scene classification. *Transactions on Audio Speech Language Processing*, *23*(1), 142–153.

Yang, S., & Chong, X. (2021). Study on feature extraction technology of real-time video acquisition based on deep CNN. *Multimedia Tools and Applications*, *80*(25), 33937–33950. doi:10.1007/s11042-021-11417-7

Zhang, D., & Wang, X. (2022). Optimization of vocal singing training methods based on multimedia data analysis. *Mathematical Problems in Engineering*, *2022*, 1–10. doi:10.1155/2022/7609516

Zhao, D., & Liu, Y. (2022). A multimodal model for college English teaching using text and image feature extraction. *Computational Intelligence and Neuroscience*, e3601545. Bokiev, D., Aralas, D., Ismail, L., & Othman, M. (2018). Utilizing music and songs to promote student engagement in ESL classrooms. *International Journal of Academic Research in Business & Social Sciences*.

Zong, Y., & Huang, G. (2021). A feature dimension reduction technology for predicting DDoS intrusion behavior in multimedia internet of things. *Multimedia Tools and Applications*, *80*(15), 22671–22684. doi:10.1007/s11042-019-7591-7

Weijuan Nie was born in Henan, China, in 1984. From 2003 to 2007, she studied in Henan Normal University and received her bachelor's degree in 2007. From 2007 to 2010, she studied in Henan Normal University and received her Master's degree in 2010. Currently, she works in Zhengzhou Preschool Teachers College. Her research interests are included Musicology.

Wan Ng is an Associate Professor at the School of Education, University of Technology Sydney, Australia.