2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom)
and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)

Visual Background Recommendation for Dance Performances Using Dancer-Shared Images

Jiqing Wen, Xiaopeng Li, James She, Soochang Park, and Ming Cheung

HKUST-NIE Social Media Lab, Hong Kong University of Science & Technology, Hong Kong
{jwenab, xlibo, eejames, eewinter, and cpming}@ust.hk

***Abstract*—Dance performances use body gestures as a language to express emotion, and lighting and background images on the stage to create the scene and atmosphere. In conventional dance performances, the background images are usually selected or designed by professional stage designers according to the theme and the style of the dance. In new media dance performances, the stage effects are usually generated by media editing software. Selecting or producing a dance background is quite troublesome, and is generally carried out by skilled technicians. The goal of the research reported in this paper is to ease this process, meaning dancers can set background images for their dance performances without the need for stage designers. Instead of searching for background images from the sea of available resources, dancers are recommended images they are more likely to use. This paper proposes the idea of a novel system to recommend images based on content-based social computing. A model to predict a dancer's interests in candidate images through social platforms, e.g., Pinterest, is proposed. With the help of such a system, dancers can select from the recommended images and set them as the backgrounds of their dance performances through a media editor. To the best of our knowledge, this would be the first dance background recommendation system for dance performances.**

***Index Terms*—dance background, dance style, image recommendation, image content**

I. Introduction

Dance is an art form in which performers use their body language to express emotions, such as affection, and the dance performance usually combines the art of dance with stage effects. There are many kinds of stage effects, such as music, lighting, smoke, and, most important, stage background. In a dance performance, the stage background helps to illustrate a story or produce a scene. For example, Fig. 1 is a photo of the ballet dance *Sleeping Beauty*. In this photo, the two dancers are in the foreground and a stage background that tells the time and place of this dance is behind them. The choice of stage background image is very important as it must be in accordance with the dance style and the artistic aesthetic.

In recent years, with the development of new media dance, some digital interactive systems have been produced to create interactive dances. A general user-friendly digital interactive system consists of three parts: 1) a motion capture device, used to track the dancer's motion and transfer motion data to the system; 2) a media editor, software for dancers to design and edit media effects; and 3) stage effect generation, where various stage effects are edited in the media editor, triggered by the dancer's motion data, and are finally displayed by the actuators, such as the projector and LED screen.

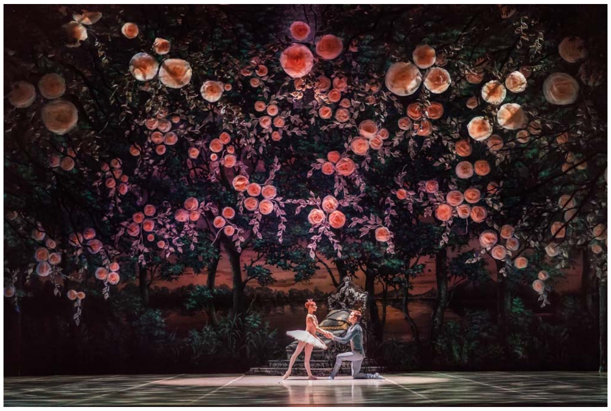

Fig. 1. The components of a dancing image: the dancer(s) and the background image.

The problem for the users of such systems, especially amateur dancers, is that they generally cannot afford a stage designer to design stage background images for them, nor can they use the available multimedia software to do media production because they lack the expertise to use it. Users must independently search for images which can be used as a dance background when holding a personal dance performance, which may lead to the problem of information overload. Thus, a dance background image recommendation system that can be used in a digital interactive system to support a media editor is required. The architecture of an advanced digital interactive system that incorporates such a recommendation system into the media editor can be seen in Fig. 2. With the help of such a digital interactive system, dancers, including amateur dancers, could easily create dance performances on their own.

To build a recommendation system, three major steps are needed, as shown in Fig. 4. The first step is to predict the top K images that the dancer (i.e., the system user) would most likely be interested in. The selected images consist of two parts: the dancer(s) and the background image; however, our final goal is to recommend images with pure stage backgrounds to system users. Hence, the second step is to get the corresponding images with pure background, i.e., remove the dancer(s) from the images. The last step is to recommend the images obtained in the second step to the dancer. This paper mainly focuses

978-1-5090-5880-8/16 $31.00 © 2016 IEEE
DOI 10.1109/iThings-GreenCom-CPSCom-SmartData.2016.120

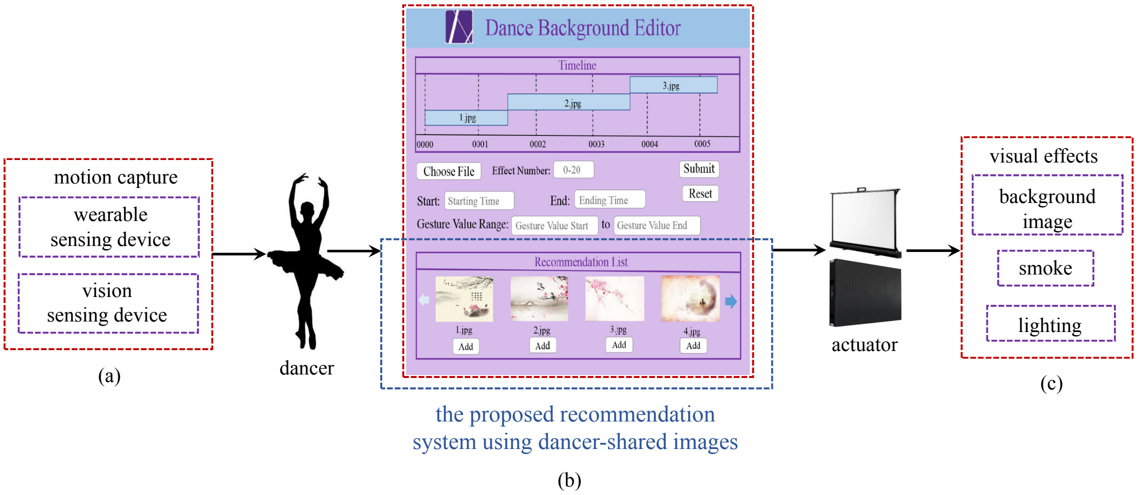

Fig. 2. The architecture of an advanced digital interactive system that incorporates a recommendation system into the media editor: (a) motion capture devices; (b) a media editor; and (c) stage effect generation.

on implementing the first step of the recommendation system, namely, predicting the top K candidate images that the dancer might have the most interest in. The remaining steps to complete the system are left for future work.

It is likely that when a dancer prepares a dance performance, he/she will refer to what he/she has already seen, liked or used before when searching the internet for stage background images. Because those images that interest the dancer represent what kind of background images he/she would like, based on his/her own dance style, for example ballet, he/she would probably use those background images that are more related to that style and the story of the dance. For example, as shown in Fig. 3, dancer A and dancer B's past dance photos reveal that they are ballet dancers since their dance performance images present ballet. Dancer C, on the other hand, is probably a street dancer, as evidenced by his dancing images. Even if two dancers share the same dance style, their dancing images may show a significant difference since the dance stories they are trying to convey are different, as shown by a comparison of dancer A and B's images in the figure. The dancers background image preferences are revealed through their past dancing images. This inspires us to predict dancers' interests based on the visual content of their dancing images shared on social networks or kept in their own collections. Another recommendation method is collaborative filtering (CF) method in which users' preferences are learned by the ratings given by the users to each item. CF method is not suitable for the background recommendation in this paper since dancers mainly share their own dancing images on social networks, and rarely have collaborative information.

The study in this paper is based on a similarity hypothesis that the images related to the same dance style tend to have similar visual content. Based on the visual content of dancing images, the dance style preference of dancers can be learned and background image recommendation can be made. The dancers can then select from the recommended images according to the storylines of the dances they are going to perform.

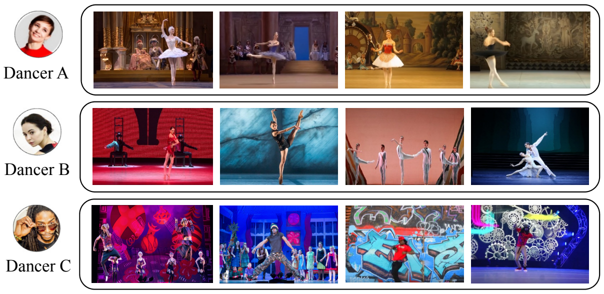

Fig. 3. From the dancing images collected by three dancers, it can be inferred that dancer A and dancer B are ballet dancers, and dancer C is mostly a street dancer.

This paper has three main contributions: 1) It proposes the idea of a novel dance background image recommendation system. To the best of our knowledge, this would be the first recommendation system that focuses on dance background images. 2) It proves a similarity hypothesis that the images related to the same dance style tend to have similar content. 3) It proposes a model to predict a dancer's interests in candidate images based on content-based social computing.

This paper is organized as follows. In Section II, related works are discussed. Section III studies the similarity of images of different styles of dancing, and the similarity hypothesis is proven. Section IV describes the design decisions of the proposed prediction methodology, while in Section V, experimental results comparing our proposed content-based method with another two methods are presented, proving the feasibility of the proposed prediction method. Finally, the conclusion and future work are given in Section VI.

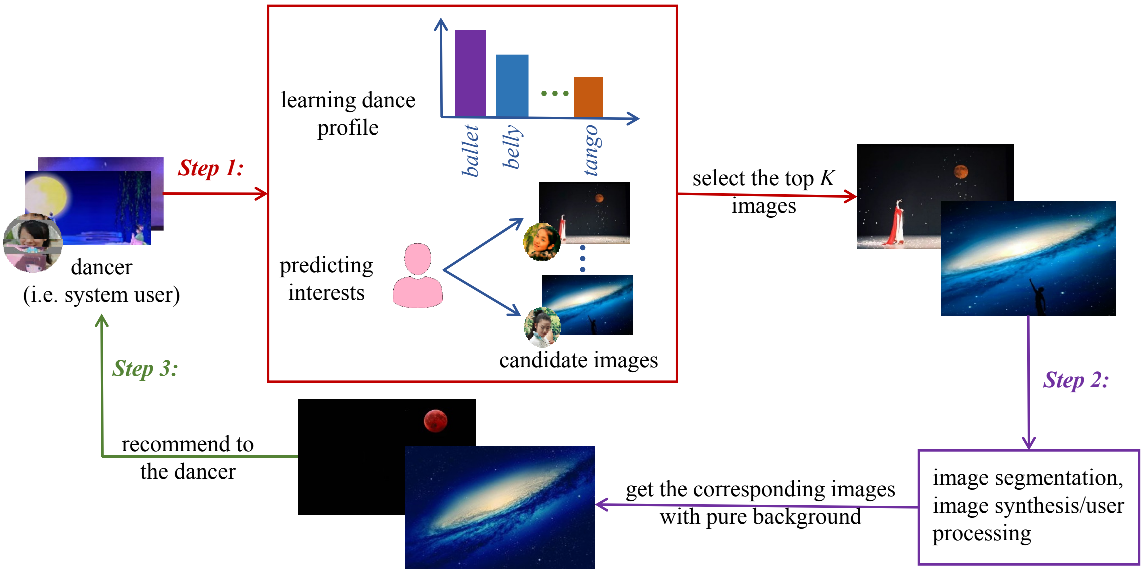

Fig. 4. The architecture of the proposed dance background image recommendation system.

II. Related works

Dance is a narrative art, and stage background images usualy help to illustrate a story or produce a scene. The choice of stage background images is very important as the images must fit in with the overall style as well as the dance content. In conventional dance performances, such as, one of the most classic ballets *Swan Lake*[1], or the famous dance in China *Lotus Heart*[2], the stage background images are selected or designed by stage designers. New media dance is a new genre of dance, which came into being in the early period of the 20th century and has developed rapidly in recent decades. There now exist some digital interactive systems which integrate motion capture, gesture recognition and media production, and many interactive dances have been created based on these systems, for example, *Lucidity* by James et al. [1] and the *Dance. Draw* project by Latulipe et al. [2], which is a project that consists of a series of interactive dance performances. As mentioned previously, the stage backgrounds in these types of works are usually produced by media editing software.

The appearance of the first papers on recommendation systems can be dated back to the mid-1990s. In this information age, people have access to plenty of information; however, they also spend a lot of time filtering the information to extract what is useful. Therefore, recommendation systems have become increasingly important. Current applications include the recommendation of products on shopping websites, and of movies, music, and news. Adomavicius and Tuzhilin in [3] give a comprehensive summary of the three main categories of how recommendations are made in such systems: content-based recommendation, collaborative recommendation, and a hybrid approach. The paper lists the methods under these three recommendation categories, analyzes their problems, and summarizes the possible extensions of recommendation system.

In content-based image recommendation, the user profile can be learned from the visual content of the images posted by a user [4]. The same idea can also be used in music recommendation [5]. The biggest problem with content-based recommendation is over-specification, which is when a user only rates a limited number of images so the content information is limited for discovering the user's preferences. Different from content-based recommendation, collaborative recommendation mainly focuses on user behavior data, such as likes, comments, or viewing histories [6]. In general, collaborative recommendation outperforms content-based recommendation; however, as described in [5], collaborative recommendation suffers from the problem of cold start, i.e., it is only applicable when usage data is available. Both content-based recommendation and collaborative recommendation are preference-based recommendation approaches. A major drawback of this kind of recommendation is that it ignores the important point that a trusted friend of a user in social networks may influence his/her behavior. To overcome this, a probabilistic model that incorporates social network information into a traditional preference-based recommendation method is developed in the work by Chaney et al. [7].

III. Similarity Hypothesis of Dancing Images in Different Dance Styles

The study in this paper is based on a similarity hypothesis that states that images related to the same dance style tend to have similar visual content. There are many different kinds of dance styles around the world, Chrisomalis summarized 105

[1] https://www.youtube.com/watch?v=9rJoB7y6Ncs
[2] https://www.youtube.com/watch?v=Rr66gbiqKt4

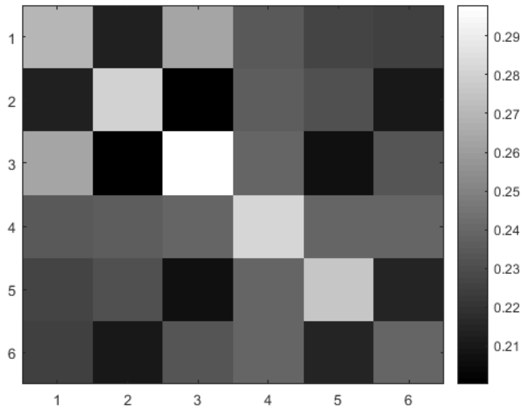

Fig. 5. Similarity of the dancing images of six styles, from 1 to 6: ballet, belly dancing, street dance, modern dance, tango, waltz.

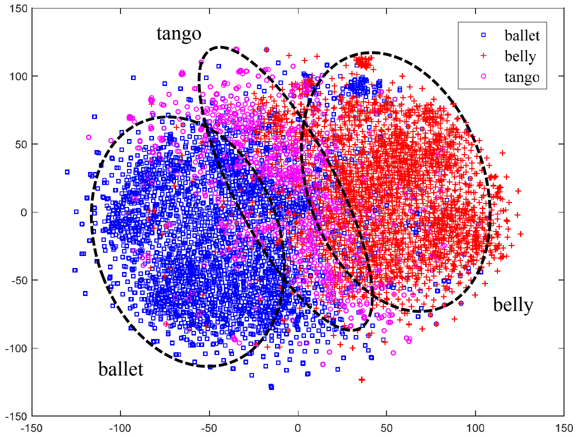

Fig. 6. Feature visualization of dancing images of three dance styles: ballet, belly dancing, and tango.

of them[3]. Dance styles vary immensely both by time period and by region, so while the summary of Chrisomalis may not be comprehensive, it provides a reference for us to recognize and understand the differences between certain styles.

We conduct an experiment in which data related to six common dance styles, namely ballet, belly dancing, street dance, modern dance, tango, and waltz, are taken from the social media network Pinterest to prove the hypothesis. Pinterest is one of the most popular image-centric social networks. On Pinterest, *pins* are images posted by users, each of which is along with a short description, and *boards* are the collections of those pins. In this experiment, 10 boards belonging to 10 different users on Pinterest are scraped for each dance style. The total number of tested pins is over 12,000. The image features are extracted using convolutional neural networks (CNNs). (The mechanism and advantages of CNNs will be described in Section 4.1.) Fig. 5 shows a confusion matrix to describe the pairwise similarities of the six dance styles, with Table I referencing the style ID and style name. The matrix element in the ith row and jth column $C(i,j)$ computes the cosine similarity of the dancing images in the ith dance style and the dancing images in the jth dance style:

$$C(i,j) = \frac{1}{NM} \sum_{n=1}^{N} \sum_{m=1}^{M} cosine(I_n, J_m), \tag{1}$$

where N and M are the total numbers of images in the ith and jth dance style, respectively. I_n and J_m denote the nth image in the ith dance style and the mth image in the jth dance style. A lighter color indicates that these two dance styles are more similar.

TABLE I
REFERENCE OF STYLE ID AND STYLE NAME

Style ID	Style Name
1	Ballet dance
2	Belly dance
3	Street dance
4	Modern dance
5	Tango
6	Waltz

To make the results more intuitive, three dance styles are selected, ballet, belly dancing, and tango, to show the feature visualization of the dancing images in these three styles. We run the t-SNE algorithm [8] to find a 2-dimensional embedding of the high-dimensional feature space and plot them as points, with the color determined by the dance style they belong to. By intuition, it should be possible to distinguish the plotted image points that are in the same dance style and those that are not. The visualization is shown in Fig. 6. The figure shows that the image points of the same dance style tend to converge into a cluster, and the distances between the image points of the same dance style are smaller than those of the image points of different dance styles, which also qualitatively proves the hypothesis.

The confusion matrix and the feature visualization, prove that dancing images of the same dance style tend to be visually more similar than those of different dance styles. This gives us the opportunity to recommend dance background images based on image visual content. Inferring the expected dance style from a dancer's collection of images or from the dancer's specifications, the system can recommend background images matching the dancer's style through analyzing the image content. However, a dancer often has an interest in a variety of dance styles. The dancer's interests can thus be modeled proportionately over all dance styles. The visual differentiability between different dance styles enables us to model a dancer's interests through visual image content. Thus, it is reasonable to recommend dance background images based on the dancer's interests and the dance style(s) reflected in the recommended images.

[3] http://phrontistery.info/dance.html

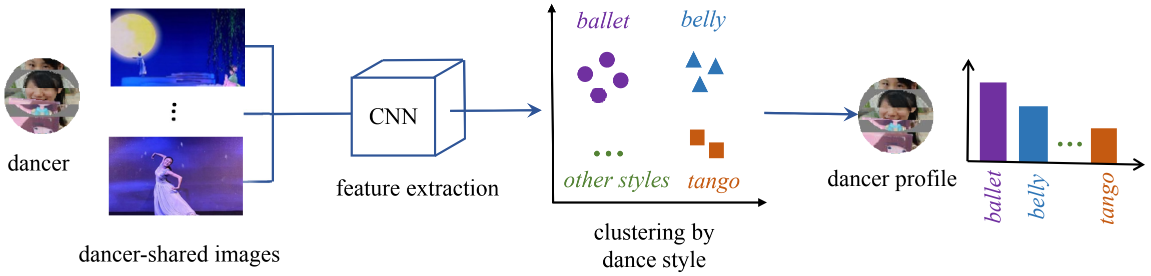

Fig. 7. Feature extraction and learning dancer profile.

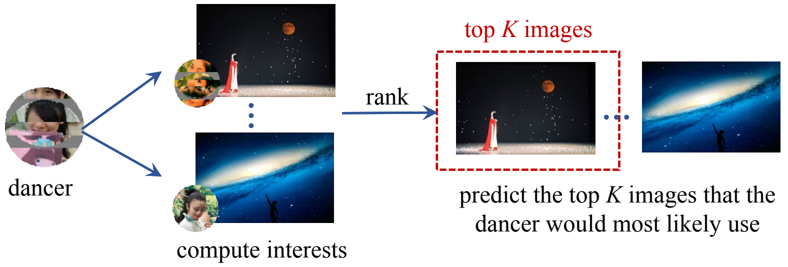
dancer
compute interests
rank
top K images
predict the top K images that the dancer would most likely use

Fig. 8. Predicting a dancer's interests in candidate images.

IV. Proposed Prediction Methodology

This paper proposes a prediction model to predict whether a user would like a given image based on the dancing images shared on social networks or in his/her collection. The goal of this section is to predict the top K images that the dancer (i.e., the system user) would most likely use as a dance background. To achieve this, there are three main procedures: 1) extracting the features of the dancing images shared by the dancer on the social network Pinterest; 2) learning the dancer's profile according to the features of the shared dancing images; and 3) predicting the dancer's interest in each candidate image and ranking the candidate images according to the dancer's predicted interests. In the following, the design decisions related to these three procedures are discussed.

A. Feature Extraction

Any dancer who has ever shared images under the *dance* category on Pinterest can be denoted by d_i, $d_i = \{d_1, d_2, ... d_{N_d}\}$, and N_d is the total number of dancers. The pins belonging to dancer i are denoted by $P^i = \{I^i_1, I^i_2, ..., I^i_{|P^i|}\}$. Given the dancer-shared images, the first step is feature extraction to obtain information from the images. This step is crucial in that if the representation of images does not capture enough meaningful information, the subsequent algorithm can not precisely characterize the dancers and thus cannot correctly discover the possible connections between them. To serve this task, this paper proposes to use a deep learning method, namely, CNNs.

In recent years, deep learning structures, in particular CNNs, have achieved an astonishing breakthrough in the computer vision area. CNNs seek to learn hierarchical features through a feature learning process with a layered architecture, where the abstraction level is increasing from a lower layer to a higher layer. Generally, two types of layers are included in the architecture, the convolutional layer and the fully-connected layer, where each layer has multiple unit nodes, called neurons, in the neural network. The proposed approach, if not specifically specified, is to use AlexNet [9] pre-trained on a large-scale image dataset, ImageNet, as the center feature extraction engine and extract features from the 7th layer, FC7, obtaining a feature vector with the dimension 4,096. The feature vector for each image I^i_j is denoted by v^i_j. As is shown in the previous section, the features extracted in this way already show evident clusters throughout the dance types.

B. Learning a Dancer's Profile

The intuition behind learning a dancer's profile is to generate a feature vector from the dancing images shared by the dancer to describe the dancer's preferences. The proportion of images of a particular dance style in the dancer's image collection reveals how much the dancer prefers that kind of dance style. For example, a ballet dancer would probably have many ballet images, which obviously shows his/her preference for ballet. Explicitly stating the dancers' preference for different dance styles, however, not only requires automated classification over all dance styles but also loses the in-class differentiation of images, i.e., the users' preferences for the images of the same dance style. Instead, the profile of each dancer d_i is derived as the mean value of the feature vectors of P^i:

$$w_i = \frac{1}{|P^i|} \sum_{j=1}^{|P^i|} v^i_j, \qquad (2)$$

where w_i is a 4096-dimensional feature vector of dancer d_i. The whole process is illustrated in Fig. 7, in which the feature vectors of the images of different dancer styles are denoted by geometric figures in 2-dimensional space. According to the hypothesis illustrated in Section III, these images are automatically clustered by dance style.

C. Predicting a Dancer's Interest in Candidate Images

As shown in Fig. 8, to predict the top K images that the dancer would most likely be interested in, the first task is to predict the dancer's interest in each candidate image. The candidate images used in this paper are the images shared by all of the dancers in the dataset, instead of only those shared by the user him/herself. The reason is that the users of the recommendation system may be professional dancers or amateur dancers, and some dancers, especially amateur dancers, do not have many dancing images themselves. Thus, it would be beneficial for them to be recommended dancing images from other dancers. Moreover, the reason for using dancing images shared by dancers, instead of directly using paintings or scene photos, is that, to support an image's suitability as a dance background image, it is imperative to use images that are known to be suitable for this purpose, i.e., they have already been used by other dancers as background images. Paintings or scene photos are not necessarily appropriate.

Given a dancer d_i and a candidate image I_n^m, the interest of dancer d_i in the candidate image can be defined as

$$S(i) = cosine(w_i, v_n^m). \tag{3}$$

After obtaining the dancer's interests in the candidate images, the images are ranked by the dancer's interests, and an image with higher dancer interest should be ranked in front of an image with lower dancer interest. Then, the top K images are selected and wait to be processed and recommended to the dancer.

V. Experiments

A. Dataset and Setup

In order to evaluate the effectiveness and performance of proposed method, a dataset is scraped from Pinterest since it is one of the most popular image-centric social networks. On Pinterest, users post *pins* (typically an image along with a short description) and organize them in self-defined *boards*, each of which is associated with one of 34 predefined categories. To ensure the reliability of the data, this paper refers to the two criteria mentioned in [10]: 1) the board should contain no less than 100 pins to guarantee that there is enough data for each user; and 2) the board should have at least one *pin* posted recently to ensure that the user is still active. After filtering, the dataset contains over 437,000 pins from 808 *dance* boards, each belongs to a different dancer. Among the images, a variety of dance styles are covered. The images are split into two groups: 20% are used as testing images, while the remaining 80% are used to do training.

The 808 dancers in our dataset are regarded as the users of the recommendation system. Firstly, the profiles of the 808 dancers can be learned by equation 2 using training images. Then, each dancer's interests in the testing images can be predicted by equation 3. For each dancer, the top K images that he/she has the most interest in can be regarded as the images the dancer would likely use as a dance background. Among the 20% testing images, only a few are actual images shared by the dancer. Those images are regarded as the groundtruth for each dancer for evaluation. Then, compute the prediction precision P by computing the percentage among the predicted images that are of the groundtruth images:

$$P = \frac{\sum_{i=1}^{N_d} h_i}{KN_d} \tag{4}$$

where N_d denotes the total number of dancers in this experiment, and h_i is the hit number of dancer i, namely, among the images recommended to dancer i, the actual number of images shared by dancer i. In the experiment, we set a series of K from 5 to 100 to see whether the different values of K will obtain a different prediction precision.

To justify the feasibility of the proposed method, the experiment compares it with two baselines, the color-based

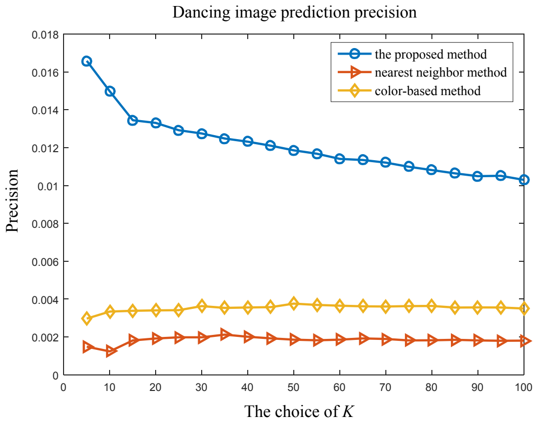

Fig. 9. The precision of the proposed method, the color-based method, and nearest neighbor method.

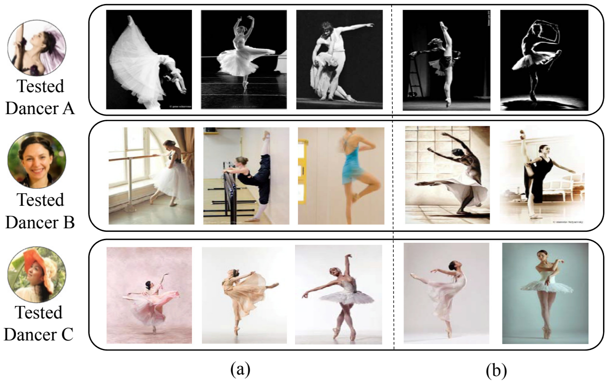

Fig. 10. Observed prediction results: (a) samples of three images shared by the tested dancer and (b) the top two images the dancer has the most interest in.

method and nearest neighbor method. We use the color-based method to justify the effectiveness of using CNNs to do feature extraction in the proposed method because it uses the same method for predicting dancer interest, but uses a color histogram to do feature extraction. In addition, to compare the effectiveness of predicting dancer interest with the proposed method, we use the nearest neighbor method, in which the image features are extracted by CNNs, and dancer's interest in each candidate image is determined by its nearest neighbor among the dancer's collection of images.

B. Results

The results of the experiment are shown in Fig. 9, where the proposed method, the color-based method, and the nearest neighbor method are plotted for comparison. As shown in Fig. 9, when using the proposed method, the prediction for dancing images works much better than the color-based method and nearest neighbor method. The highest precision reaches about 1.7%, which proves that the proposed prediction method is feasible in practice. The figure also shows that when $K \leq 15$, the precision diminishes significantly when K becomes larger, and when $K > 15$, the precision diminishes slightly when K becomes larger. In comparison, for the color-based method, due to limited representation power of color feature, it has significant inferior performance than that using CNNs. On the other hand, for the nearest neighbor method, although it also extract image content information using CNNs as the proposed method does, its ability to learning dancers' interests is worse than the proposed method. In summary, the proposed method achieves performance 6 times better than the color-based method and 4 times better than the nearest neighbor method, showing the effectiveness of the proposed method. Fig. 10 shows some examples of the recommendation results. From dancer A's shared dancing images, it tells that dancer A's interests are more in black and white ballet, and the predicted top images basically accord with his/her interests. Similar observations can be found in recommendation results for dancer B and dancer C. Qualitatively, the examples show that the proposed method indeed reveals dancers' interests through their shared images and recommend reasonable dancing images to the dancer according to their interests.

VI. Discussion

The experiment results prove the feasibility of the proposed content-based prediction method. However, in social networks, users value the opinions of their friends when it comes to discovering and discussing content, which is to say that images shared by a dancer's friends (i.e., friends in social networks) can influence the dancer's interests. Hence, future work will include improving the prediction method by combining the visual content of shared dancing images and the dancer's social information. In addition, the images selected by the proposed prediction methodology still consist of two parts: the dancer(s) and the background image; however, our final goal is to recommend images with a pure stage background. Hence, future work will also include developing another model to obtain images with a pure background.

VII. Conclusion

Stage background plays an important role in a dance performance as it helps to create the scene and atmosphere. In a conventional dance performance, the background image is usually selected or designed by professional stage designers. In a new media dance performance, the stage background is usually generated by media editing software. However, those dancers, especially amateur dancers, who cannot afford a stage designer or a media producer have to search for images which might be used as a dance background, but the overwhelming number of images available may cause a problem. Thus, this paper proposes the idea of a dance background image recommendation system for dancers. The study is based on a hypothesis that the same dance style tends to be represented by similar image content. This paper proves this hypothesis and proposes a model to predict the top K images that the dancer would most likely be interested in. In the experiment section, we compare our proposed prediction method with a nearest neighbor method and a color-based method. The method proposed in this paper shows a promising result, proving that this method is feasible in practice.

Acknowledgment

This work is supported by HKUST-NIE Social Media Lab.

References

[1] J. James, T. Ingalls, G. Qian, L. Olsen, D. Whiteley, S. Wong, and T. Rikakis, "Movement-based interactive dance performance," in *Proceedings of the 14th ACM International Conference on Multimedia*. ACM, 2006, pp. 470–480.

[2] C. Latulipe, D. Wilson, S. Huskey, B. Gonzalez, and M. Word, "Temporal integration of interactive technology in dance: Creative process impacts," in *Proceedings of the 8th ACM Conference on Creativity and Cognition*. ACM, 2011, pp. 107–116.

[3] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 6, 2005, pp. 734–749.

[4] X. Geng, H. Zhang, J. Bian, and T.-S. Chua, "Learning image and user features for recommendation in social networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4274–4282.

[5] A. Van den Oord, S. Dieleman, and B. Schrauwen, "Deep content-based music recommendation," in *Advances in Neural Information Processing Systems*, 2013, pp. 2643–2651.

[6] C. Fang, H. Jin, J. Yang, and Z. Lin, "Collaborative feature learning from social media," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 577–585.

[7] A. J. Chaney, D. M. Blei, and T. Eliassi-Rad, "A probabilistic model for using social networks in personalized item recommendation," in *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 2015, pp. 43–50.

[8] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, Vol. 9, No. Nov, 2008, pp. 2579–2605.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[10] L. Yang, C.-K. Hsieh, and D. Estrin, "Beyond classification: Latent user interests profiling from visual contents analysis," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE, 2015, pp. 1410–1416.