Classification Methods in Cultural Heritage

Marijana Ćosović[1✉][0000-0003-3398-3157], Alessia Amelio[2][0000-0002-3568-636X], Emina Junuz[3][0000-0003-1764-0576]

[1] University of East Sarajevo, Faculty of Electrical Engineering, Vuka Karadžića 30, 71126 Lukavica, East Sarajevo, Bosnia and Herzegovina
marijana.cosovic@etf.ues.rs.ba
[2] University of Calabria, DIMES, Via Pietro Bucci 44, 87036 Rende (CS), Italy
aamelio@dimes.unical.it
[3] Dzemal Bijedic University, Faculty of Information Technology, Sjeverni logor 12, 88000 Mostar, Bosnia and Herzegovina
emina@fit.ba

Abstract. This paper describes relevant classification methods applied to the cultural heritage context. In particular, a categorisation of the classification methods is provided according to tangible and intangible cultural heritage, where movable and immovable objects can be in the focus. A short description of each method is reported for each cultural heritage category in terms of feature representation, classification approach and obtained results. The proposed survey can be useful in the research community of pattern recognition and visual computing for exploring the current literature about the topic. It will hopefully provide new insights for the advancement of knowledge discovery in cultural heritage.

Keywords: Pattern recognition · Visual computing · Cultural heritage.

1 Introduction

Classification is the process of labelling data items as belonging to a given class from a model which is built from a selected set of data [1]. In particular, supervised classification aims to learn a model from a training set of data, which is then used to classify unseen test data. By contrast, the aim of the unsupervised classification is to compute the labels from the data by grouping them into meaningful classes. It can be performed using different optimisation algorithms which can make assumptions about the model of data.

Classification has been adopted in different domains, which include image processing and document analysis [1]. In image processing, supervised classification is used for identification of images, regions or pixels as belonging to a given semantic class. By contrast, unsupervised classification is used for grouping a set of images into meaningful semantic classes, or a set of pixels into homogeneous image regions according to brightness, colour, or texture. In document analysis, supervised classification can be employed for identification of documents based on authorship, typology, language, script, orthography style, dialect or sub-dialect classes. By contrast, unsupervised classification can be useful for grouping documents based on similar characteristics, e.g. language, script or similar content.

In recent time, classification in both supervised and unsupervised forms has played an important role for knowledge discovery in the cultural heritage. In particular, different classification methods for images, documents and other kinds of data have been proposed for supporting the cultural heritage understanding and preservation. Cultural heritage is an essential part of the everyday life. It includes old and contemporary pieces of art, buildings, furnitures, monuments, documents, archeological sites, as well as oral traditions and habits of different populations worldwide [9]. This provides an important motivation for exploring and analysing the different state-of-the-art classification methods in the cultural heritage domain.

In this paper, a survey of relevant works about classification methods for the cultural heritage is proposed. In particular, a categorisation of the cultural heritage is first provided, which the classification methods are included in. Then, for each category, the explored methods are described in terms of feature representation, classification algorithm and obtained results. This paper provides a useful guide for understanding classification methods in the cultural heritage domain, and evaluate their limitations. It will be useful for the proposition of innovative methods in the state-of-the-art.

The paper is organised as follows. Section 2 describes the classification methods included in the different cultural heritage categories. Finally, Section 3 draws conclusions about the presented work.

2 Cultural Heritage Methods

The term cultural heritage according to UNESCO encompasses two main categories [11]: (i) tangible, and (ii) intangible cultural heritage. Tangible cultural heritage can be categorised into: (i) movable cultural heritage (paintings, sculptures, coins, manuscripts), (ii) immovable cultural heritage (monuments, archaeological sites, historical buildings), and (iii) underwater cultural heritage (shipwrecks, underwater ruins and cities). Intangible cultural heritage has been one of UNESCO's priorities in the cultural domain recently, as it promotes cultural diversity. Through preservation of oral traditions and expressions, ways of life, traditional crafts and festivals, amongst other activities, humans are safeguarding their cultural identities.

The classification methods applied to the cultural heritage context can fit to this categorisation. Figure 1 shows the cultural heritage categorisation.

2.1 Tangible Cultural Heritage

Movable Heritage. In the document analysis context, Esposito et al. [8] proposed the use of an intelligent system for document processing based on machine

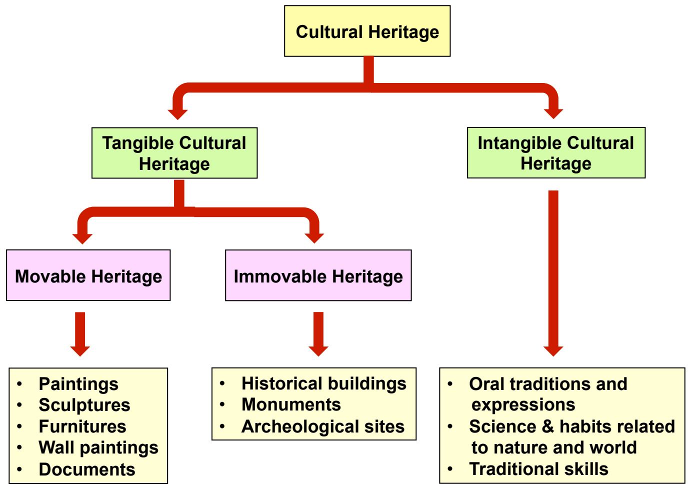

Fig. 1. Categorisation of the cultural heritage [11].

learning techniques and adapted it to the problem of automatically processing documents in film archives for the COLLATE project. The WISDOM++ system previously developed by users [8] provides a document transformation into a web-accessible form such as XML format. It is designed with high adaptivity, real-time user interaction and multi-page document management as main features. Throughout various steps of document processing, a rule base is built from a training set using machine learning techniques, hence a highly adaptive system. The authors use inductive learning techniques throughout the whole course of the document processing, consisting of document analysis, classification and understanding, and transformation into a web-accessible format. Each of these steps uses machine learning as a viable solution to various problems arising in image processing such as: classifying document components with respect to the content (separation of text from graphics), finding logical structure of the document based on layout structure extraction, text extraction from the relevant part of the document using Optical Character Recognition (OCR), and the transformation of the page into HTML/XML format. Preliminary results produced by the research are with limitations relating to the WISDOM++ system upgrade for facilitating color images, as well as images of mixed content, and further experiments are needed. Also, need for a better integration of OCR with WISDOM++ is reported to improve the usability by the end user.

Also, INTHELEX (INcremental THEory Learner from EXamples) is a learning system based on object identity assumption that learns theories from positive and negative examples and can learn multiple concepts simultaneously. It is a fully incremental learning system exploiting the induction of hierarchical logic theories examples. Also, COLLATE is a project for annotation, indexing and retrieval of digitalized historical archive material with the need to address the automatic processing of multi-format cultural heritage documents from film archives.

Considering that the authors successfully applied INTHELEX in the paper document processing domain [2], it was natural to incorporate it to COLLATE to learn rules for the automatic identification of a wide range of COLLATE document classes. The identified classes were further used for indexing and retrieval as well as annotation by users. There are 119 considered COLLATE documents, belonging to five different classes (four classes of positive and one class of negative examples). An experiment was conducted as follows: the first-order descriptions were generated by the WISDOM++ and are used to run the learning system. Scanned images were used to identify layout blocks (type and relative position) hence each document belonging to positive classes was described, on average, with 144, 215, 269, and 260 literals respectively. The features that are contained in the document descriptions are: height and width of the layout blocks, horizontal and vertical position, type of the layout and relative position. Each document is, at the same time, considered a positive example of the class it belongs to and a negative example for other classes. Definitions for each class were learnt from the initial theory that was empty. In order to compare their solution with the state-of-the-art system, the authors conducted experiments using Progol and observed accuracy and runtime of both systems confirming there are insignificant differences between the two in terms of accuracy. The runtime was significantly shorter using the INTHELEX system contributing to the incremental approach vs. batch approach in the case of Progol. Hence, the obtained results imply that the proposed solution is also a feasible solution for the automatic processing of multi-format cultural heritage documents.

In the same context, Brodić et al. [6] proposed a new document classification method for discrimination based on the language. The first phase consisted in extracting a feature representation from the document, where each letter was codified into a numerical code according to its extension in the text-line area. A total of four numerical codes were considered, corresponding to grayscale levels of an image. In this way, the document was represented by a 1-D image from which vectors of co-occurrence texture analysis features were extracted. In the second phase, the feature vectors were subjected to a genetic-based classification approach for discrimination of the corresponding documents in different languages. The algorithm represented the documents as a weighted undirected graph, where nodes were the documents and edges linked each node with its spatially close and k nearest neighbour nodes in terms of L_1 distance among the corresponding vectors. The weight on each edge corresponded to a similarity value computed from the distance between the involved nodes. Then, a genetic algorithm was applied on the graph for detecting the connected components which corresponded to groups of documents in the same language. The genetic algorithm started with a population of individuals, each randomly initialised and representing a possible division of the graph in connected components. Then, variation operators of uniform crossover and mutation were applied on each individual and the fitness function of weighted modularity was computed. This procedure was iterated for a given generation number. At the end, the individual with the best fitness function was selected as the final solution. Finally, a complete link clustering strategy for correcting the local optima was employed on the genetic solution. The experiment was performed on a database of 85 documents given in French, English, Serbian and Slovenian contemporary languages. The obtained results showed the discriminatory ability of the introduced document feature representation as well as the good performances of the genetic-based classifier versus other competing methods.

An extension of the genetic-based classifier was proposed by Brodić et al. [5] for dealing with a more complex task of discrimination of languages evolved over time. The first modification from the baseline algorithm was the introduction of a parameter in the similarity computation for reducing the sparsity of the document graph which may occur when representing close languages. The second modification consisted in managing the presence of isolated nodes of the graph inside the genetic algorithm, which can be caused by the procedure of graph construction. The same document image coding was adopted for discrimination of documents given in languages evolved one into another over time. The difference was that run-length and local binary patterns instead of co-occurrence textural features were extracted from the 1-D grayscale image of the document and combined to create the feature vector. The experiment was carried out on a database of 50 documents given in Italian Vulgar (1260-1374 AC) and modern Italian languages. Comparison results with other well-known discrimination algorithms revealed that the proposed approach overcomes the other methods as well as the baseline genetic algorithm in discriminating the different evolved languages.

In [4], the same feature coding was used with the Naive-Bayes method for identification of the orthography style of documents based on the evolution of script or language characteristics over time. In this case, the documents were stored as images, and a preprocessing phase of bounding box detection and filling was carried out for finding the position of the letters in the text-line area. This allowed the coding of the document in a sequence of numerical codes from which the grayscale 1-D image was produced. Then, run-length and adjacent local binary patterns were used on the obtained image for the extraction of the document feature vector. Finally, the Naive-Bayes classifier was employed on the vector for identification of the orthography of the corresponding document. The proposed approach was tested on two different contexts of: (i) language-based orthography identification from Serbian historical documents (classification of Slavonic-Serbian and Serbian languages), and (ii) script-based orthography identification from Croatian historical documents (classification of old and new Glagolitic scripts). Results obtained on two databases of digitised documents showed that the proposed approach is robust to data noise and overcomes other competing methods in orthography style identification.

Also, Naive-Bayes and support vector machine were adopted in [3] on the same feature coding for identification of documents as given in different Serbian pronunciations from the Shtokavian dialect. After the document image coding, different textural features were extracted from the 1-D image and combined to create the feature vector for the document. They included: (i) local binary patterns, (ii) neighbour binary patterns, and (iii) the newly introduced adjacent neighbour binary patterns. The texture operators were considered with different parameter values and tested. Then, both classifiers were used on the feature vectors for identification of the Serbian pronunciation of the corresponding documents. The experiment was performed on 50 documents written in ijekavian and ekavian pronunciations of the Shtokavian dialect in Serbian language. The obtained results showed that the proposed method overcomes the n-gram approach in the classification task.

In the furniture context, a new technique was proposed in [16] for automatic classification of archaeological pottery sherds image views contained in two databases, using features based on colour and texture. The classification task is challenging considering the pottery sherds belonging to different classes have hardly no visible differences. Ground truth images indicated by specialists are used as representatives of each class. Local features based mainly on color properties (sherds hue, chromaticity, saturation) and local texture variance of the pottery sherds are extracted for each pixel from both front and back views of the individual items. Histograms of local features are created in order to focus on prevailing information and combined with a novel bag-of-words model based on Reddi multithresholding. In this way, a global sherd descriptor for each sherd image is created. In order to reduce the global descriptors' dimensionality as well as preserve classification accuracy, the authors resorted to the Principal Component Analysis (PCA) feature selection statistical method as the most balanced amongst eight considered feature selection methods. This process offers a trade-off between the computational complexity (size of the feature matrix) and loss of information on the other side resulting in potential misclassification. The authors use several machine learning algorithms, namely K-Nearest Neighbour (KNN), support vector machine, Naive-Bayes, SMO, and Simple Logic for classification of the reduced global descriptor and have shown that the KNN algorithm is the best selection for a given method. The classification rate when using ground truth images as training example for different classes, or when splitting the dataset into training and testing parts (40% training, 60% testing) provides comparable results of about 70%. Additionally, the authors have tested their method on the ceramic sherd database as well as compared their method in both databases with four state-of-the-art feature extraction techniques (self-similarity, pyramid histogram of words–color/gray–geometric blur and a weighted combination of them) that have demonstrated an improved performance when tested in a generic image database. In the case of both tests satisfactory results were obtained.

Finally, Mensink and van Gemert [17] introduced a new large dataset composed of 112,039 artistic items from the collection of Rijksmuseum in Amsterdam, the Netherlands. The items consisted of images and associated metadata in XML format from portraits, furniture, pictures, miniatures, etc. of ancient and medieval period, and late 19th century. The dataset is open and freely available for tasks of art image classification and content-based image retrieval. Also, differently from previous datasets of specific artistic objects, such as paintings or vases, the proposed dataset is broad and variegated enough for capturing a museum-centered view. Starting from the dataset, four open challenges were proposed, including: (i) prediction of the artist, (ii) prediction of the type of artistic work, (iii) prediction of the material by which the artistic work is created, and (iv) prediction of the year of creation. For each challenge, Fisher vectors were adopted as image feature representation and support vector machine was used for classification of the artistic images. All material retrieved from the experiments is available for download, which includes dataset, experimental settings and image features.

Immovable Heritage. To protect, keep and eventually restore tangible cultural heritage buildings, systematic image collection and classification in order to correctly interpret and manage the vast information is of utmost importance. Guiding principles for recording, documentation and information management for the conservation of heritage buildings were used in [14], such that, in their works the authors created datasets comprising of ten different categories and performed classification of the images. A motivation for creating datasets of the architectural cultural heritage images was the lack of available datasets within the research community. The dataset encompasses ten different categories of Cultural Heritage images, namely Altar, Apse, Bello tower, Column, Dome (inner), Dome (outer), Flying buttress, Gargoyle (and Chimera), Stained glass, and Vault, totalling to over 10,000 images. Each image has a main element that is reflected in the labelling process. One of the authors' contributions is that the labeled images can be used as a starting point to establish superior classification techniques. They claim to have used deep learning techniques based on Convolutional Neural Network (CNN) that previously have not been used for classification of cultural heritage images and reported significant accuracy.

In [19], the authors' main motivation was the classification of three different architectural styles of buildings from images of the Mexican cultural heritage. Raw video content was used for the production of cultural heritage images, hence they contain undesirable yet not obvious elements besides the object in the focus that could impair the classification process. Machine learning techniques can address issues of identification of complex patterns in large sets of data where the objects in the focus are positioned within a specific scene (amongst other objects in the image), perhaps given in a different perspective, etc. The authors use visual attention predictors (saliency driven content selection approaches) and a traditional cropped image (centered-content) for training of a CNN in order to classify different styles of architecture. Graph based visual saliency is the model used for saliency prediction. The results show that using visual attention predictors increases the quality of the cropped data, hence better classification results are obtained based on saliency driven content selection in comparison to a traditional cropped image selection. The obtained classification rates and the authors' future work course imply that the system for batch processing of raw video footage could be devised with the contribution of creating a large number of cultural heritage images.

Also, Grilli et al. [10] aim to improve on traditional preservation and restoration methods of cultural heritage monuments. Such complex tasks require, at least, documenting and archiving architectural/historical heritage content, differentiating between various techniques of constructing buildings, and recognition of previous restoration evidence, if any. Point clouds, amongst other purposes, are used for 3D modelling of the cultural heritage monuments. Present challenges in automatic model analysis are in the domain of segmentation and classification of point clouds. Authors use 2D segmentation of the texture of 3D models generated based on three archeological case studies conducted in Italy (Villa Adriana in Tivoli; Cavea walls of the Circus Maximus in Rome; Portico in Bologna). Downsides of using this technique come from the large amount of data drawn from different types of constructing buildings as well as decorative elements used on the facades. In addition, the cultural heritage sites are from different time periods and in different stages of deterioration, hence the classification is less efficient. Supervised machine learning (based on decision tree algorithms) on UV maps, generated from the unwrapped textured 3D models, is used for classification of the 3D cultural heritage models. Accuracy and algorithm execution time were used as performance metrics. A training stage is required for supervised learning and a sufficient dataset of manually labeled images has been used. The authors propose using deep learning to get a better classification accuracy and hence propose to adequately extend the training set.

In the context of archeological sites, ICARE is an emerging digital heritage platform presently at the early stages of development [20]. It is envisioned as multimodal archiving system intended for archiving digital heritage content from potential donors and performing semantic queries for searching data archives from potential requestors. Both donors and requestors are interacting through the system that could in the future, by means of multimedia, metadata and text archives, provide believable digital cultural experiences in cultural heritage sites that are forever lost to humankind. So far the authors have collected images from cultural heritage sites in the Palmyra region and performed tests of several machine learning models for visual categorization. The bag of features framework is a novel technique for visual categorization by the support vector machine classifier that used the largest number of categories in a simultaneous experiment. Deep learning using CNN is also used by the authors for visual categorization. An episodic memory technique was adopted to achieve better classification results considering the limited image dataset. CNN using the transfer learning approach provided the best classification accuracy.

2.2 Intangible Cultural Heritage

As oral traditions, Michon et al. [18] solved the problem of identification of Arabic dialects using neural network architectures which allowed to achive satisfactory results with also a reduced number of training data for learning. In particular, the task required the discrimination of Modern Standard Arabic language and four Arabian dialects: (i) Egyptian, (ii) Gulf, (iii) Levantine, and (iv) North African, which are pretty similar. Three different runs were generated for the challenge. The first one was a neural network model characterised by a Multi-Input CNN, which separately learned lexical, acoustic and phonetic characteristics of the languages. The second run was characterised by a neural network CNN-biLSTM where the inputs were speech spectrograms, from which spatial and sequential characteristics were extracted. The last run included a binary CNN-biLSTMs. Results obtained by the different runs were compared with results obtained by a support vector machine. It was visible that Multi-Input CNN (first run) obtained the best performances versus the other methods and architectures.

Also, Ma et al. [15] explored the properties of their structural pronunciation representation for the extraction of linguistic features useful to identify the dialect or subdialect of Chinese speakers. In the first stage, the experiment involved a dialect-based speaker classification on data from 19 speakers with different dialects and sub-dialects. In the second phase, a new dataset of recordings in the different dialects and sub-dialects from an expert dialectologist was created with minimum speaker differences. Another classification task was performed on the new dataset which obtained close results than the previous experiment. Finally, a last dataset of recordings with maximum speaker differences was built. Different classification experiments were carried out on the original and simulated datasets, based on both structural and spectral comparison. Results showed that the structural comparison is able to capture the linguistic features and classification performances do not depend on the speaker characteristics.

In the traditional skills and habits context, Liu et al. [13] proposed a classification approach for identification of music pieces as belonging to cultural styles. They were represented by four different types of features: (i) timbre, (ii) wavelet coefficients, (iii) rhythm, and (iv) characteristics based on musicology. In particular, the timbral texture was adopted, which was composed of spectral shape and contrast. Also, the rhytm was characterised by strength, regularity and tempo. As musicology-based features, the chromogram, chord distribution, contrast, and pitch interval histogram were adopted. Finally, the first three moments (mean, variance and skewness) of the normalised histogram of wavelet coefficients in each level were extracted. The experiment was carried out on a dataset of 1300 music pieces, each represented by the aforementioned feature sets. Three classifiers were then employed on the dataset: (i) decision tree, (ii) KNN and (iii) multi-class support vector machine. The obtained results showed that support vector machine and KNN overcome the decision tree in classification of the music pieces in six cultural styles: (i) Western classical music, (ii) Chinese traditional music, (iii) Japanese traditional music, (iv) Indian classical music, (v) Arabic folk music, and (vi) African folk music. In particular, support vector machine obtained the highest accuracy of more than 86% when all features were used.

Also, Dimitropoulos et al. [7] presented the i-Treasures Project for management, extraction and analysis of intangible cultural heritage, which included traditional songs, dance, pottery and contemporary music pieces. In particular, the project aimed to build a platform for open access to intangible cultural heritage, transmission and exchange of knowledge. This was not only focused on digitization of cultural resources, but also on generation of new knowledge by

Table 1. Overview of the main classification algorithms (reference to the papers is reported) in the different cultural heritage categories

Classification algorithm	Tangible		Intangible
	Movable	Immovable	
rule-based	[2], [8]		
genetic-based	[5], [6]		
Naive-Bayes	[3], [4], [16]		
support vector machine	[3], [16], [17]	[20]	[13]
KNN	[16]		[13]
SMO	[16]		
Simple Logic	[16]		
CNN		[14], [19], [20]	[18]
decision tree		[10]	[13]
CRF-GMM			[12]

exploiting new multisensory-based techniques for exploring the digital content. The first step in the project was the semantic analysis of the cultural content in order to create a unified knowledge platform. It was accomplished by extracting semantics for capturing hidden patterns and relationships among elements of intangible cultural heritage, which are useful for tracking its evolution over the different generations. Image and signal processing methods were employed on the different signals for the feature extraction. Finally, a learning environment was developed using 3D technology of web-based game engines.

Finally, Li et al. [12] proposed a method for classification of traditional Chinese folk songs according to the different regional styles. A temporal model was adopted for capturing the temporal structures of the folk songs. In particular, Conditional Random Fields (CRF) were used for creating the model, whose state and transfer functions were computed by the Gaussian Mixture Model (GMM) approach. The experiment involved a dataset of 334 folk songs from the Chinese regions of Shaanxi, Jiangsu and Hunan. For each song category, the CRF model was learned from the training set. Then, the GMM was employed for fitting the audio frame features and estimating the corresponding label sequence for each CRF. In order to predict the category of a folk song in the test set, for each CRF corresponding to a specific category (regional style), the posterior probability of the folk song to belong to the given category was computed. The regional style class of the folk song was that corresponding to the highest computed posterior probability. The obtained results showed that the proposed classification method achieved an improvement of 4.6%-18.13% versus the other competing methods, i.e. support vector machine, KNN, etc.

3 Conclusion

This paper presented a categorisation of classification methods according to different types of cultural heritage. Hence, tangible and intangible heritage classification methods were described. In the tangible heritage, movable and immovable heritage classification methods were analysed. Each method was characterised by feature representation, classification algorithm and obtained results. Apart from the different features, the classification algorithms for the tangible movable heritage included: (i) rule-based algorithms, (ii) genetic algorithms, (iii) Naive-Bayes, (iv) support vector machine, and (v) KNN. By contrast, CNN and decision trees were predominantly used for tangible immovable heritage. Finally, CNN, decision trees, KNN, CFR-GMM and support vector machine were adopted for intangible heritage. Table 1 shows an overview of the described categorisation of classification algorithms. This work can be useful for analysis of the current literature in the field and for the proposition of new methods overcoming the limitations of the state-of-the-art approaches.

References

1. Amelio, L., Amelio, A.: Classification Methods in Image Analysis with a Special Focus on Medical Analytics, pp. 31–69. Springer International Publishing (2019). https://doi.org/10.1007/978-3-319-94030-4_3
2. Basile, T.M.A., Ferilli, S., Di Mauro, N., Esposito, F.: Incremental induction of classification rules for cultural heritage documents. In: Proceedings of the 17th International Conference on Innovations in Applied Artificial Intelligence. pp. 915–923. IEA/AIE'2004, Springer Springer Verlag Inc (2004). https://doi.org/10.1007/b97304
3. Brodić, D., Amelio, A.: Discrimination of different serbian pronunciations from shtokavian dialect. Procedia Computer Science **112**, 1935 – 1944 (2017). https://doi.org/https://doi.org/10.1016/j.procs.2017.08.047, knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 21st International Conference, KES-20176-8 September 2017, Marseille, France
4. Brodić, D., Amelio, A.: Recognizing the orthography changes for identifying the temporal origin on the example of the balkan historical documents. Neural Computing and Applications (Nov 2017). https://doi.org/10.1007/s00521-017-3292-1
5. Brodić, D., Amelio, A., Milivojević, Z.N.: Clustering documents in evolving languages by image texture analysis. Applied Intelligence **46**(4), 916–933 (Jun 2017). https://doi.org/10.1007/s10489-016-0878-8
6. Brodić, D., Amelio, A., Milivojević, Z.N.: Language discrimination by texture analysis of the image corresponding to the text. Neural Computing and Applications **29**(6), 151–172 (Mar 2018). https://doi.org/10.1007/s00521-016-2527-x
7. Dimitropoulos, K., Manitsaris, S., Tsalakanidou, F., Nikolopoulos, S., Denby, B., Kork, S.A., Crevier-Buchman, L., Pillot-Loiseau, C., Adda-Decker, M., Dupont, S., Tilmanne, J., Ott, M., Alivizatou, M., Yilmaz, E., Hadjileontiadis, L., Charisis, V., Deroo, O., Manitsaris, A., Kompatsiaris, I., Grammalidis, N.: Capturing the intangible an introduction to the i-treasures project. In: 2014 International Conference on Computer Vision Theory and Applications (VISAPP). vol. 2, pp. 773–781 (Jan 2014)
8. Esposito, F., Malerba, D., Semeraro, G., Ferilli, S., Altamura, O., Basile, T.M.A., Berardi, M., Ceci, M., Mauro, N.D.: Machine learning methods for automatically processing historical documents: from paper acquisition to xml transformation. In: First International Workshop on Document Image Analysis for Libraries, 2004. Proceedings. pp. 328–335 (Jan 2004). https://doi.org/10.1109/DIAL.2004.1263262

9. European-Union: The european year of cultural heritage 2018, https://europa.eu/cultural-heritage/about
10. Grilli, E., Dininno, D., Petrucci, G., Remondino, F.: From 2d to 3d supervised segmentation and classification for cultural heritage applications. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **XLII-2**, 399–406 (2018). https://doi.org/10.5194/isprs-archives-XLII-2-399-2018
11. Kurniawan, H., Salim, A., Suhartanto, H., Hasibuan, Z.A.: E-cultural heritage and natural history framework: an integrated approach to digital preservation. In: International Conference on Telecommunication Technology and Applications. pp. 177–182. Proc .of CSIT vol.5, IACSIT Press (2011)
12. Li, J., Ding, J., Yang, X.: The regional style classification of chinese folk songs based on gmm-crf model. In: Proceedings of the 9th International Conference on Computer and Automation Engineering. pp. 66–72. ICCAE '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3057039.3057069
13. Liu, Y., Xiang, Q., Wang, Y., Cai, L.: Cultural style based music classification of audio signals. In: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 57–60 (April 2009). https://doi.org/10.1109/ICASSP.2009.4959519
14. Llamas, J., M. Lerones, P., Medina, R., Zalama, E., Gmez-Garca-Bermejo, J.: Classification of architectural heritage images using deep learning techniques. Applied Sciences **7**(10) (2017). https://doi.org/10.3390/app7100992
15. Ma, X., Xu, R., Minematsu, N., Qiao, Y., Hirose, K., Li, A.: Dialect-based speaker classification using speaker-invariant dialect features. In: 2010 7th International Symposium on Chinese Spoken Language Processing. pp. 171–176 (Nov 2010). https://doi.org/10.1109/ISCSLP.2010.5684491
16. Makridis, M., Daras, P.: Automatic classification of archaeological pottery sherds. J. Comput. Cult. Herit. **5**(4), 15:1–15:21 (Jan 2013). https://doi.org/10.1145/2399180.2399183
17. Mensink, T., van Gemert, J.: The rijksmuseum challenge: Museum-centered visual recognition. In: Proceedings of International Conference on Multimedia Retrieval. pp. 451:451–451:454. ICMR '14, ACM, New York, NY, USA (2014). https://doi.org/10.1145/2578726.2578791
18. Michon, E., Pham, M.Q., Crego, J., Senellart, J.: Neural network architectures for arabic dialect identification. In: Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018). pp. 128–136. Association for Computational Linguistics (2018)
19. Obeso, A.M., Vázquez, M.S.G., Acosta, A.A.R., Benois-Pineau, J.: Connoisseur: Classification of styles of mexican architectural heritage with deep learning and visual attention prediction. In: Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing. pp. 16:1–16:7. CBMI '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3095713.3095730
20. Yasser, A.M., Clawson, K., Bowerman, C.: Saving cultural heritage with digital make-believe: Machine learning and digital techniques to the rescue. In: Proceedings of the 31st British Computer Society Human Computer Interaction Conference. pp. 97:1–97:5. HCI '17, BCS Learning & Development Ltd., Swindon, UK (2017). https://doi.org/10.14236/ewic/HCI2017.97