A Bag of Notes Approach to Writer Identification in Old Handwritten Musical Scores

Albert Gordo
Computer Vision Center
Dept. of Computer Science
Campus UAB, Ed. O, 08193
Bellatera (Barcelona), Spain
agordo@cvc.uab.es

Alicia Fornés
Computer Vision Center
Dept. of Computer Science
Campus UAB, Ed. O, 08193
Bellatera (Barcelona), Spain
afornes@cvc.uab.es

Ernest Valveny
Computer Vision Center
Dept. of Computer Science
Campus UAB, Ed. O, 08193
Bellatera (Barcelona), Spain
Ernest.Valveny@cvc.uab.es

Josep Lladós
Computer Vision Center
Dept. of Computer Science
Campus UAB, Ed. O, 08193
Bellatera (Barcelona), Spain
Josep.Llados@cvc.uab.es

ABSTRACT

Determining the authorship of a document, namely writer identification, can be an important source of information for document categorization. Contrary to text documents, the identification of the writer of graphical documents is still a challenge. In this paper we present a robust approach for writer identification in a particular kind of graphical documents, old music scores. This approach adapts the bag of visual terms method for coping with graphic documents. The identification is performed only using the graphical music notation. For this purpose, we generate a graphic vocabulary without recognizing any music symbols, and consequently, avoiding the difficulties in the recognition of hand-drawn symbols in old and degraded documents. The proposed method has been tested on a database of old music scores from the 17th to 19th centuries, achieving very high identification rates.

Categories and Subject Descriptors

I.7.5 [**Document and Text Processing**]: Document Capture—*Graphics recognition and interpretation*

General Terms

Algorithms

Keywords

Writer identification, Handwritten music scores, Bag of words

1. INTRODUCTION

In a broad sense, document categorization can be defined as the process of assigning one category to a given input document image. This categorization can be performed depending on different visual cues, such as layout configuration, global visual appearance, detection of some specific elements such as logos, seals or particular symbols. Writer identification can be another important cue for document categorization. In this case, classification is performed based on the authorship of the document. This is particulary true in the context of the analysis of historical documents. In the last years, there has been a growing interest in this area, with the purpose of the preservation, access and indexation of this artistic, cultural and technical heritage. Here, a challenging application where writer identification plays a crucial role is the retrieval of anonymous documents, and the validation of the authorship of some documents.

Most of the research on writer identification has focused on handwritten text documents. The literature is prolific in noteworthy contributions [18, 19, 20] with very good results. However, in some cases, writer identification cannot be done based on text, but on some kind of graphical information. This is the case, for example, of music scores. Since there is an important amount of old music scores without information about the composer, a writer identification approach could help musicologists in the task of identification, which is time consuming and prone to errors. In this context, the handwriting style of the hand-drawn music symbols can be used for determining the authorship of a music score. It must be said that, although some compositions contain lyrics (for singers), the aim of our work is to use only music notation because of the following reasons: first of all, it has been shown that the writer of the music symbols is not always the same writer of the lyrics, secondly, our approach will be useful for the identification of the writer in all kind of music scores, including the music scores for instruments (without lyrics).

Although some writer identification approaches [1, 10] used for logographic languages (such as the Japanese or Hebrew alphabet) make use of graphic recognition methods, few works exist on pure graphic documents. As far as we know, very few works have been performed about writer identification in old music scores. In [2, 9] a complex method was proposed, but the work was in a theoretical stage and no quantitative results were published. In [6, 7] we presented two different writer identification approaches for old handwritten music scores, inspired on some writer identification methods applied to text documents. The first method extracts fea-

DAS '10, June 9-11, 2010, Boston, MA, USA

tures for every music line, whereas the second one extracts textural features from music textures. The experimental results showed that although both methods achieved quite good identification rates (73% and 76% respectively), they are not accurate enough for a reliable writer identification.

In the current paper we propose an alternative approach, taking some ideas from the visual categorization domain, such as the *bag of visual terms* framework. Thus, the identification of the writer of a graphic document (such as an old music score) is performed by the generation of a vocabulary for graphic languages. However, visual vocabularies usually employed in visual categorization must be adapted in order to cope with the classification of graphical documents. In this paper we propose how to adapt the different steps of the generic *bag of visual terms* framework (particularly feature extraction and vocabulary construction) to generate graphical words to deal with the task of writer identification of music scores. These graphical words will be obtained without recognizing any music symbol, and consequently, avoiding the difficulties in the recognition of hand-drawn symbols in old and degraded documents. Thus, the method will be faster and more robust, as shown in the experimental results obtained on a dataset of 200 music sheets of 20 writers.

The rest of the paper is organized as follows. The Bag of Notes approach is described in Section 2, explaining the four steps involved in it: the feature extraction, the vocabulary construction, histogram representation and categorization. Experimental results are presented and discussed in Section 3. Finally, Section 4 concludes the paper and proposes some future work.

2. THE BAG OF NOTES APPROACH

The method proposed for writer identification in musical scores is based on the bag of keypoints or visual terms introduced by Gabriela *et al*. in [4] for image categorization. This method was also analogous to learning methods using the bag-of-words representation for text categorization [12, 21].

As defined in [4], *the bag of keypoints method is based on vector quantization of affine invariant descriptors of image patches.* Relevant patches are found through the images and described using a fixed length descriptor or a combination of them. These features can then be clustered into relevant *keywords* or *keypoints*, establishing a vocabulary. Finally, images can be described as a function of these vocabulary keypoints for the classification.

Analogously to the bag of visual terms, our bag of notes will consist of four sequential stages: Feature detection and description, vocabulary construction, histogram representation, and image categorization. Before these stages are carried out, a preprocessing step will be performed over the images to binarize and remove the staff lines and lyrics.

2.1 Preprocessing

The preprocessing step consists in binarizing the image and removing the staff lines and lyrics. First of all, the input gray-level scanned image is binarized with the Niblack adaptive binarization technique [16], and, in order to remove noise, filtering and morphological operations are applied. Afterwards, the music score is deskewed using the Hough Transform method, and each staff is individually rotated if necessary.

Since most of the music sheets of our database contain printed staff lines, they should be removed from the music score because there are not useful for the writer identification task. The extraction of staff lines is difficult because of paper degradation, distortions, gaps and the warping effect (see an example in Fig. 1(a)). For that reason, the following robust approach has been proposed: Firstly, a coarse staff approximation is obtained using horizontal runs as seeds to detect a segment of every staff line. This approximation is computed by applying median filters (with a horizontal mask) to the skeleton of the image. Then, the remaining horizontally-shaped symbols (see Fig. 1(b)) are used to reconstruct the staff lines. For this purpose, each segment is discarded or joined with others according to its orientation, distance and area (see Fig. 1(c)). Then, a contour tracking process is performed following the best fitting path according to a given direction. In order to cope with gaps in staff lines and to avoid deviations (wrong paths) in the contour tracking process, the coarse staff approximation above described is consulted. Afterwards, those segments that belong to the staff lines (their width is similar to the average of the width of staff lines, which has been previously computed) are removed (see Fig. 1(d)). For further details, see [6].

Finally, lyrics must be removed from the image. Although text-symbol separation can be an extremely difficult problem (e.g. when text and symbols are touching and overlapping), and thus, an intensive research should be done, it is out of the scope of our work. We have used the following hypothesis: each connected component which is not touching a staff line, will be labeled as lyrics and removed from the image. Notice that this hypothesis is valid in most cases, but it is not always true (see Fig. 1(d)). For this reason, the resulting image must be supervised in order to correct any music symbols wrongly removed, and also, removing any text that is touching the staff.

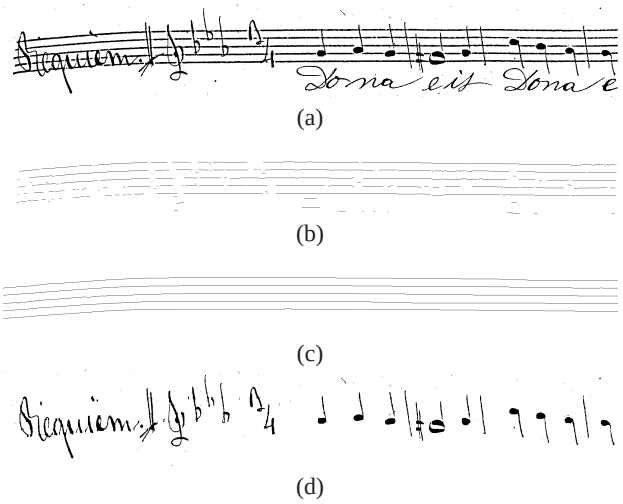


Figure 1: (a) Original Image; (b)Horizontal segments of the score; (c) Reconstruction of the hypothetical staff lines. (d) Image without staff lines nor lyrics. Notice that the word *Requiem* must be manually removed.

2.2 Feature Detection and Description

Once we have obtained the music scores without staff lines and lyrics (Fig. 2), the bag of notes approach can be performed. The first step consists in the detection of *interesting* points that will be used for describing the score image. In the image categorization domain, it is usually necessary to explore the whole image to obtain

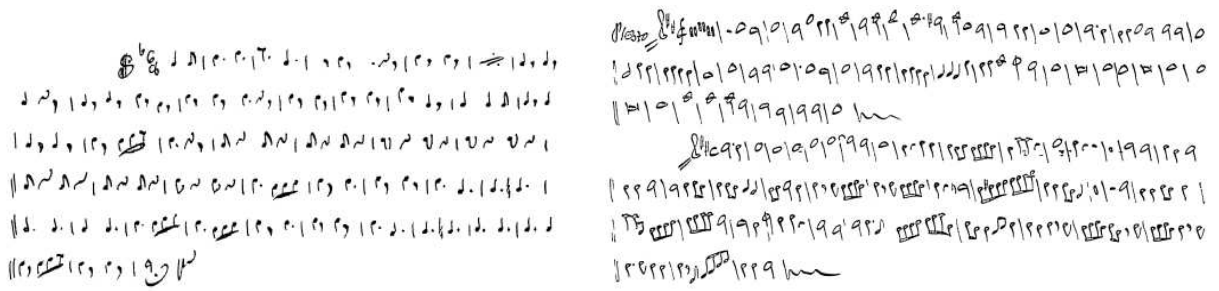


Figure 2: Samples of writing styles of two different authors after preprocessing. Colours have been inverted for printer friendliness.

these interesting points, based on edge detection, colour changes, *etc*. Common techniques for feature detection in the bag of words framework include gridding the image [22], the *Harris affine detector* [15] or Lowe's SIFT detector and descriptor [14]. Then, those interesting points are represented with one or more descriptors such as SIFT. Due to the variability of natural images, it is convenient that these descriptors are invariant to scale, rotation, illumination, *etc*. In the case of musical scores, however, we can exploit some information unavailable in other domains. Since we know that music scores are based on symbols, we can use those symbols as interesting points. After the preprocessing step just referred, we can consider each remaining element as a relevant feature, including not only music symbols (*e.g.* clefs, notes, accidentals) but also isolated graphical primitives (such as headnotes, beams, stems, *etc*).

Since we will be dealing with symbols, it makes sense to use descriptors designed for them. SIFT descriptor [14] is quite common in the image categorization domain, as it conveys some convenient features as orientation and scale invariance. However, the orientation and scale of the hand-drawn symbols provide useful information to characterize a handwriting style. Moreover, Escalera *et al*. have demonstrated in [5] that the *Blurred Shape Model* (BSM) descriptor is more suitable than SIFT for the description and recognition of hand-drawn symbols. The BSM descriptor [5] encodes the probability of pixel densities of image regions, in which each shape point contributes to a density measure of its bin and its neighboring ones (see Fig. 3). The experimental results show that BSM also outperforms other common descriptors such as Zoning, CSS or Zernike moments. Since our preliminary results agree with this analysis, we have chosen the BSM descriptor to compute the feature vectors.

2.3 Vocabulary construction

Vocabulary construction is a critical step because the final description of the images is made based on the vocabulary words. In this step we cluster the feature descriptors found in the previous step and find cluster center representatives that will form our vocabulary.

Clustering in the original bag of visual terms is done using k-means. However, more elaborate techniques can be used. In [11], Farquhar *et al*. introduce the use of generative models as *Gaussian Mixture Models* (GMM) in the vocabulary construction, where this generative information can later be exploited in the histogram representation.

Vocabulary construction can also be performed in an unsupervised or supervised way. Unsupervised clustering can be performed over the features of all the classes, yielding an *universal* vocabulary. An alternative consists in performing supervised clustering over all the classes separately and then combining the results, yielding *adapted* vocabularies. Unsupervised clustering is simpler than the supervised one, but it is computationally more expensive as the working space is much bigger. Other caveats include difficulties to represent particularities of each class or the need to recalculate if a new class is added. Supervised clustering, albeit slightly more complex, is usually faster as the quantity of features to cluster each time is much lower. It can deal with the particularities of each class and does not need to be recalculated when adding new classes. However, it is not exempt of problems. For example, supervised clustering is not exploiting common similarities between classes and can in fact cause problems with redundant clusters during the histogram representation.

For our experiments, clustering will be performed by means of a GMM, and both unsupervised and supervised approaches will be tried and compared.

2.4 Histogram representation

After a vocabulary has been built, it is necessary to represent each image as a function of it. If we only have information about the center means, as in k-means clustering, this is reduced to label the features of the image as their closest vocabulary point and build an histogram of them.

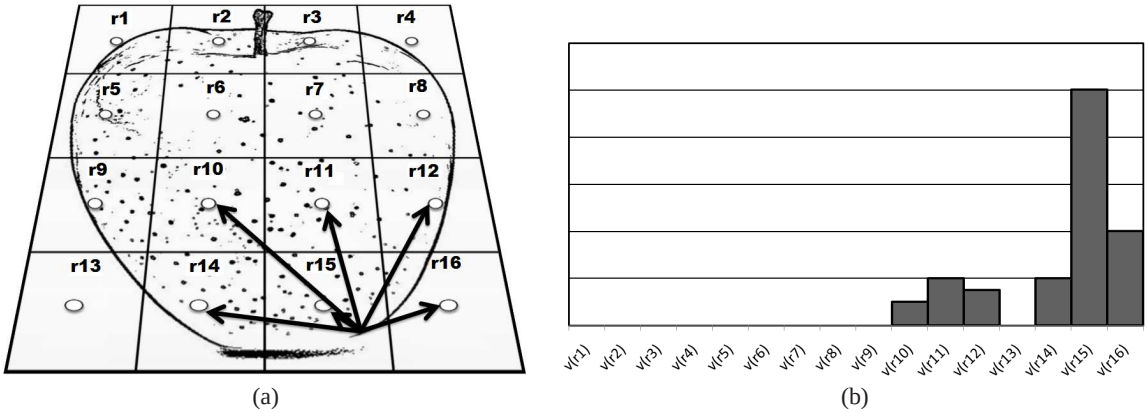
Given a set of T low level features of D dimensions obtained from an image $X = \{x_t, t = 1 \dots T\}$, and given the centers of the k-means $\mu = \{\mu_n, n = 1 \dots N\}$, where N is the number of vocabulary words, we can define function C as:

$$C_t(i) = \begin{cases} 1 & \text{if } i = \underset{n}{\operatorname{argmin}} \sqrt{(\mu_n - x_t)'(\mu_n - x_t)}, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

And finally, the relative number of appearances of the i-th word in an image will be given by:

$$\frac{1}{T} \sum_{t=1}^{T} C_t(i). \quad (2)$$

The problem with this approach is that each feature is assigned to one and only one vocabulary word, when sometimes the distinction is not so clear. This is particularly true in the case of supervised clustering, as two vocabulary words from different classes may be



(a) (b)

Figure 3: BSM density estimation example (extracted from [5])

essentially equal, but only one of them can be chosen for each feature. A softer quantization of the vector can be achieved if instead of performing a hard assignment, we use the distances to all the cluster centers to represent the feature. This still has two problems: it assumes that all the clusters are spherical and that all the weights are the same.

In the case of clustering using GMMs, we have more information and we can overcome such problems, building the histogram based on the posterior probabilities of the features in the GMM.

Let us define a Gaussian Mixture Model $\lambda = \{w_i, \mu_i, \Sigma_i, i = 1 \dots N\}$ where w_i, μ_i and Σ_i represent the weight, mean vector and covariance matrix of Gaussian i, where $\sum_{i=1}^{N} w_i = 1$, and where N is the number of Gaussians. We will also assume that Σ_i is a diagonal covariance matrix. Then, each Gaussian represents a word of the visual vocabulary, w_i represents the relative frequency of word i, μ_i the mean of the word and Σ_i the variation around the mean.

Then, the probability of x given λ will be

$$p_i(x|\lambda) = \frac{exp\{-\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)\}}{(2\pi)^{D/2}|\Sigma_i|^{1/2}}, \quad (3)$$

and the probability of feature x_t being generated by the i-th Gaussian will be:

$$\gamma_t(i) = p(i|x_t, \lambda) = \frac{w_i p_i(x_t|\lambda)}{\sum_{j=1}^{N} w_j p_j(x_t|\lambda)}. \quad (4)$$

Finally, the relative number of appearances of the i-th word in an image will be given by:

$$\frac{1}{T}\sum_{t=1}^{T}\gamma_t(i). \quad (5)$$

In this case, the feature affects in a weighted way that all the words of the vocabulary and not only the closest one, solving the previously exposed issue.

In our particular case, the use of hard assignment would present yet another problem, particularly when combined with an unsupervised learning of the vocabulary. As a plain Bag of Words just represents the frequency of the words, this would be equivalent to counting the frequency of quarter notes, eight notes, *etc*, leading to a rhythm based representation of the scores. The use of soft histograms combined with supervised clustering should alleviate this problem, as a particular symbol will likely affect a set of clusters and not just its closest one. Moreover, variations of the bag of words framework allow to go beyond counting, *c.f*. Section 4, Conclusions and Future Work.

2.5 Categorization

Finally, once the images have been described based on the vocabulary words, the problem is reduced to a multi-class supervised classification. In [4] both SVM and Naïve Bayes classifiers are used. Results in this case show a much better performance of the SVM over the Naïve Bayes classifier. In [17], however, a Sparse Logistic Regression [13] is used instead of SVM with similar results.

Since both SVM and SLR classifiers obtain very similar results in related problems, we will use a free implementation of the SVM classifier.

3. EXPERIMENTS

3.1 Dataset

We have tested our approach in a data set consisting of 200 music sheets, containing 10 pages for each one of 20 different writers. These pages are extracted from a collection of spanish old music scores of the 17th, 18th and 19th centuries, which have been obtained from the archive of the Seminar of Barcelona and the archive of Canet de Mar. A sample of the scores can be seen at Fig. 4. The music sheets have been scanned using a flatbed scanner, and stored in bitmap format. They have been captured in gray-scale at a resolution of 300 dpi, which is enough for capturing the information contained in the image.

3.2 Experimental setup

For the sake of comparability with [8], we have also used 5-fold cross validation, choosing one page per writer for each test subsets and averaging the results. Following that methodology, we will also test our approach with subsets of 5, 10, 15 and 20 writers.

For feature representation, BSM size has been experimentally fixed at 8×8, obtaining a good trade-off between the classification results and the descriptor size (64 elements). The vocabulary clustering will be performed by means of a GMM, both unsupervised (sizes 16, 32, 64, and 128 Gaussians) and supervised (combination of 2, 4, 8, 16 and 32 Gaussians per class). Features will be described

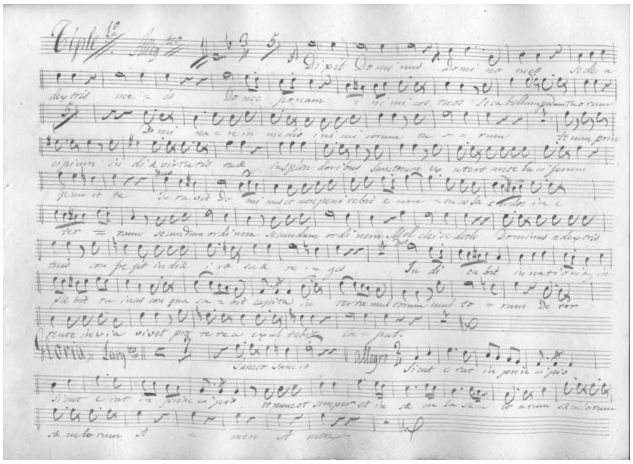


Figure 4: Example of an old score of the composer Clausell.

Table 1: Writer identification accuracy with unsupervised clustering

		Clusters			
		16	32	64	128
Number of writers	5	96	**100**	**100**	**100**
	10	94	**100**	98	**100**
	15	90.6	93.3	96	**97.3**
	20	90	94	**96**	**96**

based on their posterior probabilities over each of the Gaussians in the vocabulary.

Classification is performed with the LIBSVM [3] implementation of a SVM classifier. A radial basis function kernel will be used, where the parameters have again been experimentally fixed to $C = 15$ and $\gamma = 15$. It should be noted that variations of these parameters do not offer significant variations in the results.

3.3 Results

Results of the experiments with unsupervised and supervised clustering can be seen at tables 1 and 2, respectively. We can see that, for the best setup in each case, the results are essentially equal in both unsupervised and supervised approaches, with differences no higher than 1%. In the case of unsupervised clustering, best results are obtained with 64 and 128 Gaussians. With supervised clustering, best results seem to be obtained when using between 16 and 32 Gaussians per class, yielding 320 or 640 words in the case of 20 writers. Notice how this is a low number of words compared to image categorization problems, where the number of words is usually in the order of thousands.

Finally, Fig. 5 shows the best results for our unsupervised and supervised approaches compared to those shown in [8]. We can see that not only the results obtained here are better in every case, but also that the scalability is better. It is reasonable to think that increasing the number of writers would yield even bigger differences between methods.

Table 2: Writer identification accuracy with supervised clustering

		Clusters per class				
		2	4	8	16	32
Number of writers	5	**100**	96	**100**	**100**	**100**
	10	90	98	**100**	**100**	**100**
	15	90.6	94.6	**97.3**	**97.3**	**97.3**
	20	93	92	95	96	**97**

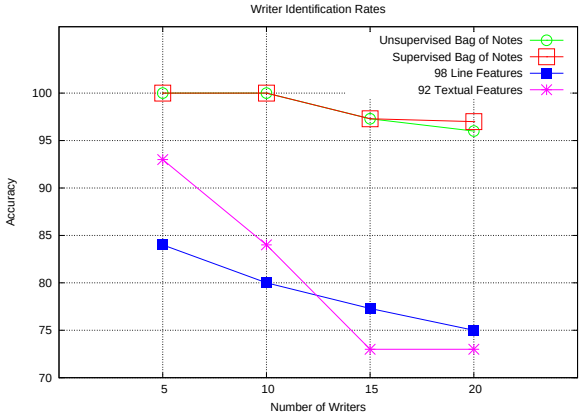


Figure 5: Comparison of Bag of Notes and methods in [8] accuracy as a function of the number of writers.

3.4 Discussion

In the natural images domain, clusters of features do not usually carry any semantic meaning, as interesting points have been chosen based simply on some features as edges or corners, colour variations, *etc*. On the other hand, in this particular scenario of music score classifications, the interesting points we selected are symbols of the score, and we could assume the vocabulary centers obtained in the clustering do have a semantic meaning, *e.g.*, one center would represent whole notes, another would represent variations of a treble clef, *etc*. However, such "high level" information clustering would be problematic. For example, the quarter notes shown in Fig. 6 belong to different authors, but, unless enough clusters have been defined, most of them will end up in the same cluster. In the supervised case, where each cluster was trained over one particular author, notes from such author will likely have an important weight in that cluster, so this is not necessarily an issue. On the other hand, in the case of unsupervised clustering, this could yield a symbol dependent representation of the score, which would not be helpful for a writer identification task. We are more interested in a clustering based on information about graphical primitives, such as stem thickness, flag shapes, head roundness, *etc*.

Unfortunately, when performing unsupervised clustering, and particularly with a low number of words, clustering will be based on the overall shape of the symbols and not on their details, forming clusters of whole notes, half notes, clefs, *etc*, regardless of the writer. With this kind of clustering, writer identification rates should definitely be much lower than the results we obtain (90% with as few as 16 clusters), and certainly deserve an analysis.

A deeper inspection of the scores and the clusters reveals an inter-

Figure 6: Quarter notes of different authors.

esting fact: most of the notes have been "broken", and so their head and stem have been considered as different components. There are mainly two reasons for this. First, it can be part of the author's style to draw the stem slightly separated from the head of the note. Another option is that the preprocessing is more aggressive than it should and breaks some notes. This is not unreasonable in the staff removal stage of the preprocessing.

In this case, clusters are mostly made around note heads and stems, providing an advantageous situation. Clustering is no longer based on higher level symbols as notes (except on the cases where notes have not been broken) but features concerning graphical primitives (e.g. headnotes, stems, flags). This avoids the problem of having to represent all or most of the high level symbols (think, *e.g.*, in upward and downward stem notes) that now can be represented as a combination of these features from graphical primitives. Even if not all the details can be accurately represented with such low number of words, the combination of features from graphical primitives adds discriminant information that whole symbols do not contain. Moreover, note that the bounding box of unbroken notes contain a lot of white space, usually in the same locations, losing discriminatory power. However, when describing note heads and stems, empty space is more significant since it characterizes the writer's style.

This should immediately raise one question: should not all notes be broken in such this way as part of the preprocessing stage? The answer is not so clear. First of all, since the separation of note head and stem is after all a writer discriminant feature, it is not completely clear whether actively breaking them would benefit the classification when using a different configuration than unsupervised clustering with few words. The second reason is that such an action would require some kind of note detection and recognition stages during the preprocessing. This is certainly possible, but one of the advantages of our bag of notes approach in respect to previous methods is precisely that we do not need to perform any kind of early classification. In Section 4, conclusions and future work, we will discuss a possible approach to this problem that does not need to break the notes exploiting the generative information of the GMM using Fisher Kernels [17].

4. CONCLUSSIONS AND FUTURE WORK

In this paper, we have proposed a new approach to writer identification in handwritten musical scores that does not require any kind of symbol recognition. Instead of that, a bag of visual terms approach is followed, where the musical score components are described and clustered with no need to recognize each music symbol.

Experimental results show that this approach outperforms state-of-the-art methods with 5, 10, 15 and 20 writers, and the evolution of the results suggests that further increasing the number of writers would yield even higher differences between methods.

However, the method still has room for improvements. Probably the most interesting would be the use of the generative information of the GMM to further improve the histogram description and go beyond counting as shown, *e.g.*, by Perronnin and Dance in [17]. The Fisher kernel can be used to obtain a representation with much more information than just the posterior probabilities of the Gaussians. Intuitively, we represent the features not just with the probabilities of each cluster but also with their "position" in it. This could solve the clustering problem when the notes are not broken. Even if *e.g.*, quarter notes of different authors end up in the same cluster, they will end up in different "positions", leading to different image descriptions. Another advantage of this Fisher representation is that the number of words needed to obtain similar or equal results severely decreases, as a consequence of each word now containing much richer information. However, we are already using very few words to begin with, so the advantages of this description beyond the broken notes problem are arguable. It should also be noted that the price of this richer information is a dramatical increase of the histogram size.

The feature description by BSM is also open to improvements. In our experiments, we have fixed its size to an 8×8 grid. It would be interesting to build a description obtained by combining several BSM resolutions. In this case, however, it would be important to apply a PCA dimensionality reduction in order to avoid typical correlation problems when combining multiple resolution descriptors.

Finally, we are interested in increasing the size of our music score database, both in number of writers and in pages per writer.

Acknowledgements

The authors would like to thank Prof. Josep Maria Gregori Cifré from Art Department of UAB for his help in accessing to old resources of archive of Seminar of Barcelona and the archive of Canet de Mar. This work has been partially supported by the Spanish projects TIN2008-04998, TIN2009-14633-C03-03 and CONSOLIDER - INGENIO 2010 (CSD2007-00018).

5. REFERENCES

[1] I. Bar-Yosef, I. Beckman, K. Kedem, and I. Dinstein. Binarization, character extraction, and writer identification of historical Hebrew calligraphy documents. *International Journal on Document Analysis and Recognition*, 9(2):89–99, 2007.

[2] I. Bruder, T. Ignatova, and L. Milewski. Integrating knowledge components for writer identification in a digital archive of historical music scores. In *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital libraries (JCDL)*, pages 397–397, New York, NY, USA, 2004. ACM.

[3] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[4] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *In Workshop on Statistical Learning in Computer Vision, ECCV*, pages 1–22, 2004.

[5] S. Escalera, A. Fornés, O. Pujol, P. Radeva, G. Sánchez, and J. Lladós. Blurred Shape Model for binary and grey-level symbol recognition. *Pattern Recognition Letters*,

30(15):1424–1433, 2009.

[6] A. Fornés, J. Lladós, G. Sánchez, and H. Bunke. Writer identification in old handwritten music scores. In *8th International Workshop on Document Analysis Systems*, pages 347–353, Nara, Japan, September 2008.

[7] A. Fornés, J. Lladós, G. Sánchez, and H. Bunke. On the use of textural features for writer identification in old handwritten music scores. In *Document Analysis and Recognition (ICDAR). Tenth International Conference on*, volume 2, pages 996–1000, July 2009.

[8] A. Fornés, J. Lladós, G. Sánchez, and H. Bunke. Symbol recognition: current advances and perspectives. In *Eight International Workshop on Graphics Recognition (GREC)*, pages 186–197, La Rochelle, France, July 2009.

[9] R. Göcke. Building a system for writer identification on handwritten music scores. In *Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA)*, pages 250–255, Rhodes, Greece, 30 June – 2 July 2003.

[10] S. Hirose, M. Yoshimura, K. Hachimura, and R. Akama. Authorship Identification of Ukiyoe by Using Rakkan Image. In *Document Analysis Systems, 2008. DAS'08. The Eighth IAPR International Workshop on*, pages 143–150, 2008.

[11] H. M. J. Farquhar, S. Szedmak and J. Shawe-Taylor. Improving bag-ofkeypoints image categorisation. Technical report, University of Southampton, 2005.

[12] T. Joachims. Text categorization with suport vector machines: Learning with many relevant features. In *ECML '98: Proceedings of the 10th European Conference on Machine Learning*, pages 137–142, London, UK, 1998. Springer-Verlag.

[13] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink. Sparse multinomial logistic regression: fast algorithms and generalization bounds. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(6):957–968, June 2005.

[14] D. G. Lowe. Object recognition from local scale-invariant features. *Computer Vision, IEEE International Conference on*, 2:1150, 1999.

[15] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *In Proceedings of the 7th European Conference on Computer Vision*, pages 0–7, 2002.

[16] W. Niblack. *An Introduction to Digital Image Processing*. Prentice Hall, 1986.

[17] F. Perronnin and C. Dance. Fisher kernels on visual vocabularies for image categorization. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, June 2007.

[18] H. Said, T. Tan, and K. Baker. Personal identification based on handwriting. *Pattern Recognition*, 33(1):149–160, January 2000.

[19] A. Schlapbach and H. Bunke. Off-line writer identification and verification using gaussian mixture models. In S. Marinai and H. Fujisawa, editors, *Machine Learning in Document Analysis and Recognition*, volume 90 of *Studies in Computational Intelligence*, pages 409–428. Springer, 2008.

[20] L. Schomaker and M. Bulacu. Automatic writer identification using connected-component contours and edge-based features of uppercase western script. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):787–798, June 2004.

[21] S. Tong and D. Koller. Support vector machine active learning with application sto text classification. In *ICML '00: Proceedings of the Seventeenth International Conference on Machine Learning*, pages 999–1006, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[22] J. Vogel and B. Schiele. Natural scene retrieval based on a semantic modeling step. In *CIVR*, pages 207–215, 2004.