Multiscale Anisotropic Texture Unsupervised Clustering for Photographic Paper

S.G. Roux[(1)], N. Tremblay[(1)], P. Borgnat[(1)], P. Abry[(1)], H. Wendt[(2)] and P. Messier[(3)]
[(1)] Laboratoire de Physique, ENS Lyon, CNRS 46 allée d'Italie, 69007, Lyon, France
[(2)]CNRS, IRIT-ENSHEEIT, Université de Toulouse, France
[(3)] Paul Messier LLC, Boston, USA.
Email: patrice.abry@ens-lyon.fr, stephane.roux@ens-lyon.fr,
Work supported by ANR BLANC 2011 AMATIS BS0101102 and GRAPHSIP ANR-14-CE27-0001-02

***Abstract*—Texture characterization of photographic papers is likely to provide scholars with valuable information regarding artistic practices. Currently, texture assessment remains mostly based on visual and manual inspections, implying long repetitive tasks prone to inter- and even intra-observer variability. Automated texture characterization and classification procedures are thus important tasks in historical studies of large databases of photographic papers, likely to provide quantitative and reproducible assessments of texture matches. Such procedures may, for instance, produce vital information on photographic prints of uncertain origins. The hyperbolic wavelet transform, because it relies on the use of different dilation factor along the horizontal and vertical axes, permits to construct robust and meaningful multiscale and anisotropic representation of textures. In the present contribution, we explore how unsupervised clustering strategies can be complemented both to assess the significance of extracted clusters and the strength of the contribution of each texture to its associated cluster. Graph based filterbank strategies are notably investigated with the aim to produce small size significant clusters. These tools are illustrated at work on a large database of about 2500 exposed and non exposed photographic papers carefully assembled and documented by the MoMA and P. Messier's foundation. Results are commented and interpreted.**

I. INTRODUCTION

In historical studies of photographic papers, surface examination constitutes an important task, likely to provide photo conservators with crucial information regarding artistic intentions, manufacturer practices, assessment of texture matches or outliers in an artist production. Often, texture assessment remains mostly achieved by visual or manual inspections. This implies long, tedious and repetitive tasks, which are moreover prone to inter- and even intra-observer variability and cannot be achieved for large size databases, cf. e.g., [1], [2], [3].

There is thus a strong need for automated procedures achieving reproducible and quantitative texture characterization. This gave birth to several recent attempts and efforts to investigate the relevance and performance of computer-assisted automated texture characterization and clustering procedures in the context of historical studies of photographic papers, cf. e.g., the *Historic Photographic Paper Classification* Challenge lead by R. Johnson and P. Messier, under the supervision of the Museum of Modern Art, New York (cf. `http://www.papertextureid.org` and [3]).

As a first step towards automation, a raking light imaging device, the *TextureScope*, has recently been designed to measure photographic paper texture with a standardized procedure, currently accepted by numerous museums, thus permitting reproducible and quantified texture measurement [1], [3]. The paper sheets analyzed in this contribution were digitized according to that procedure.

Once digitized samples are made available, image processing tools are needed to characterize textures. Several of them, of very different natures were recently compared in [3], [4]. Following our first works [5], [6], texture characterization in the present contribution makes use of the Hyperbolic Wavelet Transform [7], providing anisotropic multiscale characterization of textures [8].

Texture characterization produces a collection of features from which distance between prints, ressemblance and dissemblance can be quantified. In the context of photographic print classification, several unsupervised classification procedures were compared in [3], [4]. In previous contributions [5], [6], Spectral Clustering [9] was used. A well-known shortcoming of spectral clustering, as well as of other unsupervised clustering techniques such as K-means, lies in the fact that the number K of clusters is either arbitrarily chosen or decided via the recourse to external criteria (such as e.g., AIC, BIC) [10]. In addition, such approaches necessarily classify all images, including those consisting of potential outliers actually far from any cluster. The present contribution aims to alleviate such shortcomings by revisiting spectral clustering as being an ideal low-pass filtering on graphs and considering other multiscale lowpass filterbanks instead, following what has been done for community detection in networks [11]. This elaborate on a first attempt [12] on a larger dataset.

We illustrate how measures of cluster stability and cluster core permit to estimate the relevant number(s) of clusters as well as the relevance of a given print's attribution to a cluster.

This is applied to a large dataset of photographic papers, prepared by P. Messier, and detailed in the following Section.

II. PHOTOGRAPHIC PAPER DATASET AND FEATURES

A. Dataset

The dataset consists of 2491 images obtained using the *texturescope*, a raking light close up image acquisition that provide repeatable standardized conditions. This imaging system, extensively described in [3], requires minimal specialized handling so that large image sets can be produced rapidly. The images obtained have a size of 1536×2080 pixels and depict a surface of 1.00×1.35 cm (1 pixel equivalent to $42.2\mu\text{m}^2$). This selected scale reveals some microscopic features, such as matting agents occasionally used by manufacturers, but also depicts attributes recognizable to a human observer. The dataset includes blank and printed paper with a very large diversity of origin and material, and consists in several subsets:

- 2031 sample images of traditional black and white paper surface texture taken directly from manufacturer packages or sample books spanning the 20th century. These samples are representative of the full range of surface textures available to 20th century photographers. Documentation about texture, reflectance, brand, manufacturer and year are provided by art experts and curators for a large proportion of the dataset samples, essentially consisting of blank papers ;

- 348 sample images from Thomas Walther collection held by the Museum of Modern Art and contain work by leading modernist photographers primarily active in Central and Eastern Europe between the World Wars ;

- 83 sample images from Man Ray (1890-1976) prints ;

- 18 sample images from Lewis Hines (1874-1940) prints ;

- 11 prints belonging to the Museum of Fine Arts, Houston (MFAH) from the same artist and depicting the same images as 11 prints from the Thomas Walther collection.

B. Multiscale features

Texture characterization is achieved using the Hyperbolic wavelet transform [7], [8], a variation of the two dimensional discrete wavelet transform which relies on two independent dilation factors along the x_1 and x_2 directions. This transform explicitly takes into account the possible anisotropic nature of image textures. More precisely, the hyperbolic wavelet coefficients of an image I are obtained by scalar product of the digitalized image with a separable wavelet dilated by two different factors, a_1 and a_2:

$$T_I((a_1, a_2),(k_1, k_2)) = \frac{1}{\sqrt{a_1 a_2}} \langle I(x_1, x_2), \left(\psi(\frac{x_1 - k_1}{a_1}, \frac{x_2 - k_2}{a_2})\right)\rangle. \tag{1}$$

The multiscale representation of the image is then obtained by computing space averages (l^2–norm) of these coefficients at fixed scale (a_1, a_2):

$$S_I(a_1, a_2) = \tfrac{1}{n_a} \textstyle\sum_{k_1,k_2} |T_I(a_1, a_2, k_1, k_2)|^2, \tag{2}$$

Cepstral distance: To quantify proximity between images I and J, we use a cepstral-type distance of $S_I(a_1, a_2)$:

$$\mathscr{D}(I, J) = \textstyle\sum_{a_1,a_2} |\tilde{S}_I(a_1, a_2) - \tilde{S}_J(a_1, a_2)|, \tag{3}$$

where $\tilde{S}_I(a_1, a_2) = \log\left(S_I(a_1, a_2) \;/\; \sum_{a_1',a_2'} S_I(a_1', a_2')\right)$.

This distance does not depend upon a change in the intensity of the raking light and exposure variables that influence overall image brightness. The selected analysis scales are $1 \leq a_1, a_2 \leq 7$ and correspond to physical scales ranging from $13\mu\text{m} \leq a_1, a_2 \leq 83\text{mm}$ (7 octaves) thus yielding a matrix of 49 multiscale features. The similarity matrix is then defined using the non-linear transformation $\mathbf{W} = \exp(-\mathscr{D}/\epsilon)$ where ϵ is a constant assessing the typical closeness between images (of the order of the standard deviation of the distance between all pairs of the database). A comparison with classical clustering tools (PCA, K-means...) apply directly on $\mathscr{D}$ will be done in a future work.

III. Unsupervised clustering with graph filters

A. Background : Spectral clustering

There are many ways to cluster data in an unsupervised manner (see, e.g., [10]) and we focus in this work on the network interpretation of the similarity matrix $\mathbf{W}$ as a weighted adjacency matrix. Finding groups in the data can be tackled as finding a partition of the corresponding undirected weighted graph (each node being a data sample).

We first recall the method of spectral clustering [10], [9] to partition a graph. The normalized Laplacian matrix is $\mathscr{L} = \mathbf{I}_N - \mathbf{S}^{-\frac{1}{2}} \mathbf{W} \mathbf{S}^{-\frac{1}{2}}$, where $\mathbf{S}$ is the diagonal matrix of node strengths, with $\mathbf{S}_{ii} = \sum_{i \neq j} \mathbf{W}_{ij}$, and $\mathbf{I}_N$ is the identity matrix of size N (the number of nodes). $\mathscr{L}$ is diagonalizable, and its sorted eigenvalues are: $0 = \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_N \leq 2$; associated to orthonormal eigenvectors: $\boldsymbol{\chi} = (\boldsymbol{\chi}_1 | \boldsymbol{\chi}_2 | \ldots | \boldsymbol{\chi}_N)$. Classical spectral clustering makes use of the first K eigenvectors to define feature vectors $\boldsymbol{f}_{K,i} \in \mathbb{R}^K$ for each node i:

$$\forall k \in [1, K] \qquad \boldsymbol{f}_{K,i}(k) = \boldsymbol{\chi}_k(i). \tag{4}$$

Then, any suitable classification method can be used to obtain K clusters from these feature vectors. The K-means algorithm or hierarchical clusterings are often used.

B. Graph filters for spectral clustering

It was first proposed in [11] to use graph wavelet filters (introduced in [13]) to find relevant sub-groups that partition the data. This work was done for the mining of communities in graphs and it was later introduced in [12] in the context of unsupervised clustering. Let us consider the analogy stating that the graph Laplacian's eigenvectors are equivalent to the graph's Fourier modes [14]. It follows from this analogy that filtering on graphs is defined by a filter kernel which is diagonal on the Laplacian eigenvector basis. Following [15], we will use low-pass filters (derived from a kernel function h) designed in this graph Fourier space $[0, 2]$ as being scaling function filterbanks (in relation to wavelet filterbanks). Refer to [15] for their exact mathematical expression.

Then, a scale parameter $s \in \mathbb{R}^{+*}$ is introduced to dilate the kernel h in order to explore different levels of resolutions of the clusters to be found. Specifically, the discrete filter vector $\boldsymbol{h}_s$ reads $\forall i \in [1, N] \quad \boldsymbol{h}_s(i) = h(s\lambda_i)$, and the scaling function centered around node i at scale s has value $\boldsymbol{\phi}_{s,i} = \boldsymbol{\chi} \mathbf{H}_s \boldsymbol{\chi}^\top \delta_i$ where $\mathbf{H}_s = \text{diag}(\mathbf{h}_s)$. Given these scaling functions, we define new feature vectors different from the $\boldsymbol{f}_{K,i}$ of classical spectral clustering :

$$\boldsymbol{f}_{s,i} = \boldsymbol{\chi}^\top \boldsymbol{\phi}_{s,i} = \mathbf{H}_s \boldsymbol{\chi}^\top \delta_i. \tag{5}$$

Note that if one defines $\boldsymbol{h}_s$ as a simple low pass step function up to λ_K (i.e. $\boldsymbol{h}_s(i) = 1$ for all $\lambda_i \leq \lambda_K$ and zero otherwise), then $\boldsymbol{f}_{s,i}$ is exactly equivalent to the classical feature vectors $\boldsymbol{f}_{K,i}$ of Eq. (4). These new feature vectors may therefore be seen as a generalization of spectral clustering feature vectors to filter kernels that may be different from the step function involved in spectral clustering (as illustrated in Fig. 1).

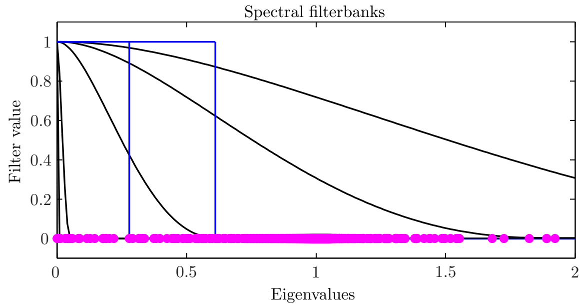

Fig. 1. Filterbanks for scaling functions (black lines) at scales $s = 1, 10, 30, 50$, for comparison against the step function filters (in blue) involved in classical spectral clustering. The magenta dots correspond to the eigenvalues of the normalized Laplacian matrix $\mathcal{L}$.

C. A stochastic estimate of the cosine distance between feature vectors

The next step is to estimate a distance. Here, we choose the cosine distance[1] between feature vectors associated to any pair of nodes i and j :

$$\boldsymbol{D}_s(i,j) = 1 - \frac{\boldsymbol{f}_{s,i}^\top \boldsymbol{f}_{s,j}}{||\boldsymbol{f}_{s,i}||_2 \; ||\boldsymbol{f}_{s,j}||_2}. \tag{6}$$

Note that :

$$\begin{aligned}\frac{\boldsymbol{f}_{s,i}^\top \boldsymbol{f}_{s,j}}{||\boldsymbol{f}_{s,i}||_2 \; ||\boldsymbol{f}_{s,j}||_2} &= \frac{\boldsymbol{f}_{s,i}^\top \boldsymbol{\chi}^\top \boldsymbol{\chi} \boldsymbol{f}_{s,j}}{\boldsymbol{f}_{s,i}^\top \boldsymbol{\chi}^\top \boldsymbol{\chi} \boldsymbol{f}_{s,i} \; \boldsymbol{f}_{s,j}^\top \boldsymbol{\chi}^\top \boldsymbol{\chi} \boldsymbol{f}_{s,j}} \\ &= \frac{\boldsymbol{\phi}_{s,i}^\top \boldsymbol{\phi}_{s,j}}{||\boldsymbol{\phi}_{s,i}||_2 \; ||\boldsymbol{\phi}_{s,j}||_2}.\end{aligned}$$

The cosine distance between $\boldsymbol{f}_{s,i}$ and $\boldsymbol{f}_{s,j}$ is equal to the cosine distance between the scaling functions associated to i and j. To compute $\boldsymbol{D}_s$, one needs first to compute all scaling functions $\boldsymbol{\phi}_{s,i}$, which requires the diagonalisation of the (potentially large) Laplacian matrix. To circumvent this issue, and inspired by Section 5 of [11] where the cosine distance between wavelets is stochastically estimated, we define stochastic features as :

$$\boldsymbol{\ell}_{s,i}^\top = \boldsymbol{\phi}_{s,i}^\top \boldsymbol{R} = \delta_i^\top \boldsymbol{\chi} \mathbf{H}_s \boldsymbol{\chi}^\top \boldsymbol{R} \tag{7}$$

where $\boldsymbol{R} = (\boldsymbol{r}_1 | \boldsymbol{r}_2 | \ldots | \boldsymbol{r}_\eta)$ is a set of η realizations of random vectors, of N i.i.d. Gaussian random variables of zero mean and finite variance. We can show, following [11] for wavelet filters, that the correlation of the feature vectors $\boldsymbol{\ell}_{s,i}$, of size η for each node, gives an estimate of the cosine distance between scaling functions :

$$\lim_{\eta \to +\infty} \widehat{\mathrm{Cor}}(\boldsymbol{\ell}_{s,i}, \boldsymbol{\ell}_{s,j}) = \frac{\boldsymbol{\phi}_{s,i}^\top \boldsymbol{\phi}_{s,j}}{||\boldsymbol{\phi}_{s,i}||_2 \; ||\boldsymbol{\phi}_{s,j}||_2} = 1 - \boldsymbol{D}_s(i,j). \tag{8}$$

Furthermore, even though the limit holds for η infinite, in practice, a small η is enough. Empirically we observe that to correctly estimate a partition in K clusters, one needs to use at least $\eta > K$ random vectors.

The advantages of the stochastic approach are two-fold. First, it is not necessary anymore to diagonalize the Laplacian because one can rely on the fast graph filtering proposed in [13] to directly compute Eq. (7) without explicitly knowing $\boldsymbol{\chi}$. This method relies on a polynomial approximation of the filter h which is extensively explained in references [16]. Second, the inherent stochasticity allows us to assess the stability of the obtained results, which in turn helps us to estimate the relevant number of clusters.

[1]This section's argumentation is true for the cosine distance. Another classically used distance is the correlation distance, which would only slightly change the following calculations.

D. Clustering algorithm

We have just shown that given η random vectors grouped in matrix $\boldsymbol{R}$, and computing the correlation of their scaling function transform, one obtains an estimate of the cosine distance matrix $\boldsymbol{D}_s$. We then use hierarchical clustering with average linkage as in [11] to obtain a partition in clusters P_s.

We generate J random matrices $\boldsymbol{R}$, and obtain thereby a family of partitions $\{P_s^j\}_{j \in [1,J]}$. Because of the randomness in the definition of $\boldsymbol{R}$, any two partitions in this family are not necessarily equal. This stochasticity allows us to propose two indices to assess whether an obtained clustering is robust enough: the global stability of the clusters, and the local constancy of the attribution of each node to cluster(s).

E. Robustness of the clustering

The first metric, called Global Stability Index $\gamma(s)$, is defined as the mean of the similarity between all pairs of partitions of $\{P_s^j\}_{j \in [1,J]}$:

$$\gamma(s) = \frac{2}{J(J-1)} \sum_{(i,j) \in [1,J]^2, i \neq j} \mathtt{ari}(P_s^i, P_s^j) \tag{9}$$

where the function $\mathtt{ari}$ is taken here as the Adjusted Rand Index [17]. The more stable is the partition at scale s, i.e. the more interesting this scale is, the closer to 1 will be $\gamma(s)$. In the following, we will closely consider scales whose stability is high enough. In fact, we typically estimate the clustering for a set of logarithmically spaced scales between two boundaries automatically detected by the algorithm [11]. Only the scales s^*, the local maxima of $\gamma(s)$, and their associated partitions are kept. This allows us to circumvent the usual issue of choosing or estimating the number of clusters K in spectral clustering.

Second, we introduce a local robustness measure of the clustering. Core clusters were defined in [18] in the context of community detection and we extend its use to the problem of data clustering. Two nodes i and j will be in the same cluster core if, in each of the J partitions, they are always classified in the same cluster. Clusters of size 1 will not be counted as cores. Given a scale s, from the J different clustering $\{P_s^j\}_{j \in [1,J]}$, we obtain a list of cluster cores $\{C_z\}_{z \in [1,Z]}$. For each node i, we define its frequency of core association $\rho_s(i)$: if i belongs to core C_z, then the frequency is $\rho_s(i) = 1$; if i is not in a core, we compute the relative frequency $\rho_s(i) = n_z / J$ based on the number of occurrences n_z for which i is associated to the same core C_z (we only keep the z' which maximizes n_z). In some cases, a is always in its own cluster of size 1 which does not qualify as a core cluster and we keep it unclassified with $\rho_s(a) = 0$. If $\rho_s(i) = 1$, then we may confidently classify data i in its cluster core. If not, then i cannot be classified with full reliability with other points in this dataset, and C_z (cluster with a maximal number of occurrences for i) is only an indication of its closest core. The average

$$\bar{\rho_s} = \frac{1}{N} \sum_a \rho_s(i) \tag{10}$$

over all nodes of their relative association frequency to a cluster core gives a second stability function that we will call Local Stability Index in the following. The closer is $\bar{\rho}_s$ to 1, the more stable is the local association of nodes (i.e., prints) to cluster cores.

Global and Local Stability Indices are illustrated in Fig. 2.

F. Choice of the Laplacian

The equivalence between spectral clustering and low pass graph filtering is technically exactly valid for symmetrical Laplacian matrices (whose eigenbases are orthonormal): the combinatorial Laplacian $\boldsymbol{L} = \boldsymbol{S} - \boldsymbol{W}$ and the normalized Laplacian (as previously defined) are two good candidates. One may extend this analogy to the random walk Laplacian $\mathscr{L}_{rw} = \mathbf{I}_N - \mathbf{S}^{-1}\mathbf{W}$, which is not symmetrical anymore but has the same spectrum and has shown better results in the spectral clustering litterature (see [9] for instance). In the following, we will show results obtained with the random walk Laplacian.

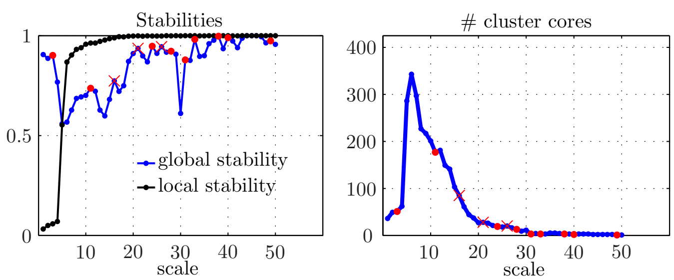

Fig. 2. Left: Global (Eq. (9), blue) and Local (Eq. (10), black) Stability Indices of achieved clustering versus clustering scale. Red dots on the Global Stability Index correspond to local maxima, and thus indicate scales at which achieved clustering show most relevance. The three red cross correspond to the local maxima of the three analyzed scales in section IV. Right: Corresponding number of cluster cores at each scale.

IV. Photographic paper clustering

A. Stability analysis and hierarchy of clusters

The proposed stochastic scaling function filterbank clustering is applied to the affinity matrix $\mathbf{W}$ defined in Section II-B and computed from features measured on each sample of the large dataset described in II-A: 50 different scales are scanned, using $\eta = 200$ and $J = 100$ random matrices.

Fig. 2 reports the obtained stability indices across scales for this large dataset (left) and the corresponding number of clusters (right). Red dots on the Global Stability Index indicate scales at which clustering can be considered relevant. While fine scale $s = 4$ is marked as stable by the Global Stability Index, the Local Stability Index is found very low at that scale indicating that most prints randomly switch from one cluster to the other in the stochastic procedure used to compute the cosine distance (cf. Section III-C). This is due to artifacts of the similarity measure Adjusted Rand Index used in the definition of γ in Eq. (9). Scale $s = 4$ is thus considered as not yielding a relevant clustering. The Global Stability Index shows 10 local maxima above scale $s > 10$. In all cases, these local maxima coincide with large Local Stability Indices. These scales can thus be considered as yielding relevant partitions of the datasets of prints at hand.

At the finest stable scale $s = 11$, the achieved partition

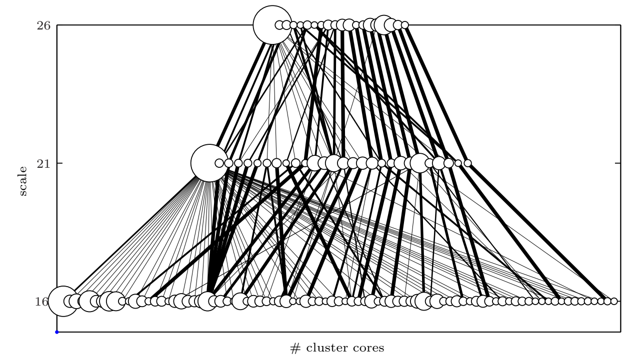

Fig. 3. Hierarchical representation of three stable scales (out of the eleven detected in Fig. 2): scales $s = 16, 21, 26$ with respectively $85, 28, 20$ clusters (from bottom to top). The width of the link connecting clusters at two different scales quantifies the proportion with which the coarse scale cluster feeds the fine scale one. The size of the dot indicates the size of the clusters. Achieved clustering is not necessarily hierarchical.

consists of 177 clusters, the largest of which contains 487 prints. The precise analysis and interpretation of the content of each cluster is beyond the scope of the present article. For the illustration of the relevance of the proposed clustering approach, let us now concentrate on three stable scales (amongst the 9 remaining stable scales), $s = 16, 21$ and 26, yielding respectively $85, 28$ and 20 clusters. Fig. 3 represents the evolution of the core communities along the three chosen scales $s = 16, 21$ and 26. It shows that the achieved clustering is not hierarchical: a cluster at scale s can split into several at scale $s' < s$, and conversely, a cluster at scale s' can results from the merging of (portions) of clusters at a coarser scale $s > s'$.

Fig. 4 displays the mean 49-features, $\tilde{S}_X(a_1, a_2)$, represented as an image, for the 20 cluster cores obtained at coarse scale $s = 26$. The differences in the mean signatures permit to assess differences between the anisotropic multiscale properties of the prints classified into each cluster: Asymmetry in amplitude around the first diagonal betrays asymmetry of the texture ; Large amplitude close to the bottom right corner indicates energy at coarse physical scales.

The proposed clustering approach relies on a true multiscale strategy, that does not favor clusters of similar size but rather clusters having the same “local vision” of the graph at a given scale. This implies that the large cluster (Cluster 1 of size 2162) consists of a large ensemble that gathers most typical prints, while the 19 other clusters have positions in the graph that are definitely peculiar at that scale, and thus correspond to prints whose properties depart significantly from the typical properties of Cluster 1. At intermediate scale $s = 21$, the large Cluster 1 splits into several clusters with one larger than the other ones (size 1944). At fine scale $s = 16$, this large cluster further splits into several clusters, the largest of which yet remains significantly larger than all others (size 830).
Let us focus on some case studies of the 19 smaller clusters at scale $s = 26$ and study with some details their evolution across scales.

B. Case studies

Silk clusters: At coarse scale $s = 26$, Clusters 13, 14 and 20 (of sizes 25, 30 and 3) contain 45 prints documented as Silk textures. Out of the 13 remainders, 9 are not documented

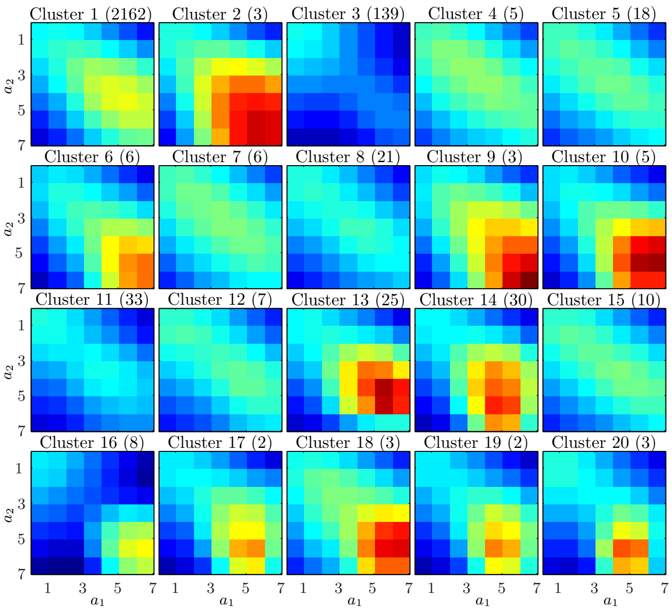

Fig. 4. Intra-cluster mean 49-features quantifying the mean anisotropic multiscale properties $\tilde{S}_X(a_1, a_2)$ of each of the 20 clusters obtained at coarse scale $s = 26$. Amplitudes are coded from blue (low) to red (high), with identical colormaps for all clusters.

and visually look very much like silk texture. Fig. 5 (left) displays one example of print and the mean features of the cluster. Obvious differences in mean features explain why the prints are clustered in 3 different groups. Fig. 3 further shows that these three clusters remain identical at intermediate scale $s = 21$. At fine scale $s = 16$, however, while Clusters 13 and 20 stay identical, Cluster 14 is split into two clusters (57 and 65, of size 11 and 20). Fig. 5 (right) illustrates differences in the mean-features of these two refined clusters. Moreover, at this fine scale, a new small cluster (77 of size 2, Fig. 5 bottom right) appears as further extraction of Silk texture papers from the very large Cluster 1 at coarse scale $s = 26$, where they were mis-classified. At this fine scale, only 6 prints documented as silk are not accounted for. They are displayed in Fig. 6 with their features. The top 4 prints show, at least visually, patterns similar to other Silk prints, with much less contrasted textures though. The 2 bottom prints, though documented as Silk, show very different textures and features. This case study of Silk textures shows the ability to extract given attributes, such as Silk texture, in an automated way, and to detect potential mislabeling in metadata.

Glossy clusters: Let us now concentrate on Clusters 3, 8 and 11 (of size 139, 21 and 33) at coarse scale $s = 26$. Metadata indicates that they all essentially contain print reflectance documented as Glossy. Fig. 7 displays a sample and mean features for each of the 3 clusters and shows that mean-features of Cluster 3 significantly depart from those of Clusters 8 and 11. Interestingly, metadata indicate that Cluster 3 preferentially consists of prints from Kodak manufacturer (64 out of 139), while Clusters 8 and 11 contain far fewer Kodak prints (2 out of 21 and 7 out of 33, respectively). The fine scale features of textures may thus be interpreted as a Kodak manufacturer signature. These clusters remain identical at intermediate scale $s = 21$. At finer scale $s = 16$, Cluster 11 remains identical, while Cluster 8 splits into three

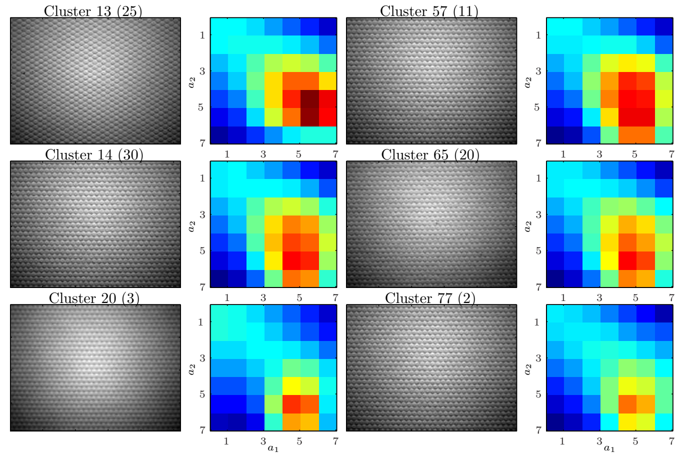

Fig. 5. Left: three clusters with texture documented as Silk at coarse scale $s = 26$. Right: at fine scale s = 16, Cluster 14 is split into two clusters (57 and 65), and a new small cluster (77) is created (Clusters 13 and 20 remain identical). For each cluster, a sample image and the mean features are represented using the same colormap.

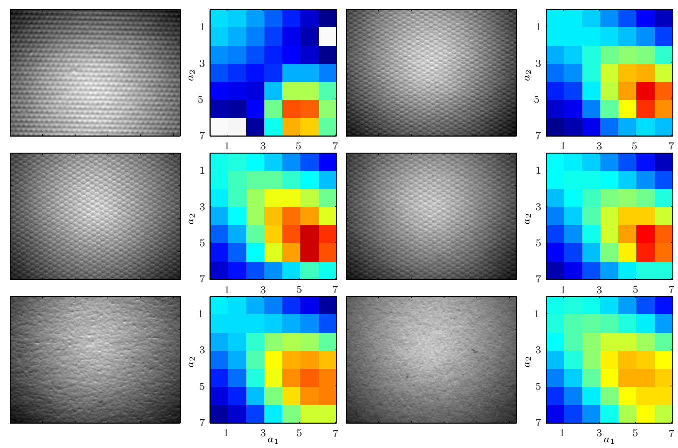

Fig. 6. The only 6 prints of the database documented as Silk that are not captured in the Silk clusters analyzed in Fig. 5 and their features. The colormap is the same as in Fig. 5.

clusters (42, 44 and 68, of size 12, 2 and 7), cf. Fig. 7 (right). Interestingly, Cluster 3 splits into four clusters shown in Fig. 8: Two of them, Clusters 53 and 56 (of size 11 and 101), contain mostly Kodak prints (8 out of 11 and 51 ou of 101) ; Cluster 18 (of size 24) contains hardly any Kodak prints (3 out of 24) ; Cluster 85 (size 2) gathers the only two prints documented as Kodak manufacturer and Scientific Imaging Film brand.

Other clusters: At coarse scale $s = 26$, Clusters 9 (size 3), 10 (size 5) and 18 (size 3) display similar features as shown in Fig 4. Cluster 10 contains the only two prints from the entire database documented as Agfa Ansco and brand Cykon or Cikora with Crystal texture. This cluster remains identical down to the finest stable scale, thus showing its consisting of a set of 2 very specific textures. Cluster 9 contains one print from the Thomas Walther collection and the only two prints documented as Defender manufacturer, Velour Black Brand with canvas texture. At fine scale $s = 16$, the two defender remain together in a cluster of size 2. Besides the large cluster (Cluster 1), the two remaining significant clusters at coarse

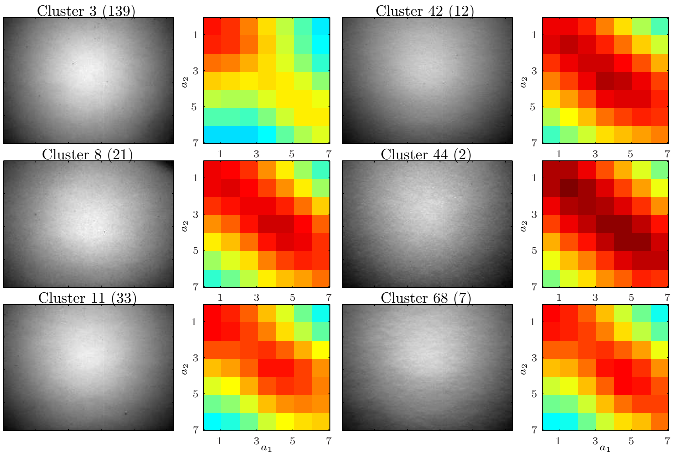

Fig. 7. Left: clusters with glossy texture at Scale 26 (20 clusters). Right: cluster 8 split in three clusters at scale 16. For each cluster we show a typical image and the mean features of the cluster using the same colormap.

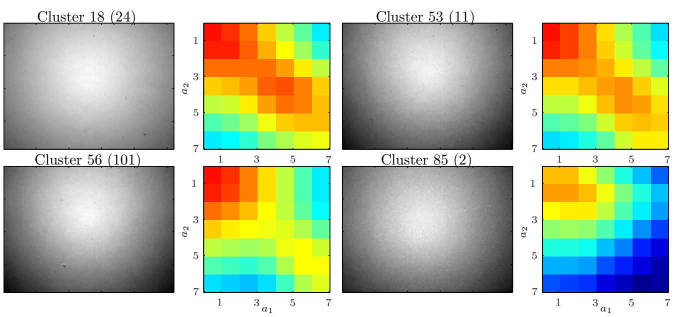

Fig. 8. Cluster number 3 at scale 26 (20 clusters) split in four clusters at scale 16 (85 clusters). For each cluster we show a typical image and the mean features of the cluster. The colormap is the same as in Fig. 7.

scale $s = 26$ (5 and 15 of size 18 and 10) do not receive simple interpretations.

V. Conclusion

This contribution elaborates on spectral clustering, relying on the framework of filterbanks in the graph Fourier domain. Instead of ideal low-pass filterbanks, use is made of low-pass scaling functions filterbanks. While any multiscale clustering technique, including the one proposed here, can output clusters with size as targeted, one of the advantages of the technique developed here lies in the indices of stability. Indeed, the stochastic evaluation of the cosine distance between graph nodes is used not only to reduce computational load, but also to construct two measures of local and global stabilities. This permits to select scales (and thus number of clusters) for which the partitioning can be regarded relevant and to assess a probability of correct classification for each node. We are thus technically able to create arbitrarily small clusters, and assess their relevance to partition data. The stability indices might drop drastically at fine scales, thus indicating that such small clusters do not make sense. For art historians, measures of the confidence in clustering may prove a great added value as it may avoid the burden of trying to interpret clusters and classifications that are actually not relevant or stable. Further, the possibility of keeping some samples unclassified may help to avoid to draw conclusions based on unreliable associations. Such algorithms and approaches may be used to achieve a number of tedious systematic tasks, such as detecting mis-labeling in documentation or proposing automatic tentative labeling by semi-supervised strategies: art historians would then only have to validate proposed pre-documentations.

On the technical side, an extension of the work may consist of exploring the replacement of low-pass with band-pass filters (wavelet filterbanks on graphs) [13]: this may yield further details at finer scales.

References

[1] C. Parker and P. Messier, "Automating art print authentication using metric learning." in *Proceedings of the Twenty-First Innovative Applications of Artificial Intelligence Conference*. Association for the Advancement of Artificial Intelligence, 2009, pp. 122–127.

[2] P. Messier, "The Paul Messier black and white photographic papers collection," 2013.

[3] C. R. Johnson, P. Messier, W. A. Sethares, A. G. Klein *et al.*, "Pursuing automated classification of historic photographic papers from raking light photomicrographs." *Journal of the American Institute for Conservation*, vol. 53, no. 3, pp. 159–170, 2014.

[4] P. Messier and C. R. Johnson, Jr., "Texture feature extraction for the classification of photographic papers," in *Proc. Asilomar Conf. Signals, Systems, Comput.*, Nov. 2014.

[5] P. Abry, S. G. Roux, H. Wendt, P. Messier *et al.*, "Multiscale anisotropic texture analysis and classification of photographic prints," *IEEE Signal Proc. Mag.*, vol. 32, no. 4, pp. 18–27, July 2015.

[6] S. G. Roux, P. Abry, H. Wendt, and B. V. S. Jaffard, "Hyperbolic wavelet transform for historic photographic paper classification challenge," in *Proc. IEEE Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, USA, Nov 2014, to appear.

[7] R. A. DeVore, S. V. Konyagin, and V. N. Temlyakov, "Hyperbolic wavelet approximation," *Constructive Approximation*, vol. 14, no. 1, pp. 1–26, 1998.

[8] S. G. Roux, M. Clausel, B. Vedel, S. Jaffard *et al.*, "Self-Similar Anisotropic Texture Analysis: The Hyperbolic Wavelet Transform Contribution," *IEEE Trans. Image Proc.*, vol. 22, no. 11, pp. 4353–4363, 2013.

[9] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.

[10] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[11] N. Tremblay and P. Borgnat, "Graph wavelets and community mining," *IEEE Trans. on Sig. Proc.*, vol. 20, no. 62, pp. 5227–5239, 2014.

[12] N. Tremblay, S. G. Roux, P. Borgnat, P. Abry *et al.*, "Texture classification of photographic papers: improving spectral clustering using filterbanks on graphs," in *Proc. GRETSI Symposium Signal and Image Processing*, Lyon, France, Sept 2015, to appear.

[13] D. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 129–150, 2011.

[14] D. Shuman, S. Narang, P. Frossard, A. Ortega *et al.*, "The emerging field of signal processing on graphs," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[15] N. Tremblay and P. Borgnat, "Multiscale community mining in networks using spectral graph wavelets," *2013 Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, 2013.

[16] D. Shuman, P. Vandergheynst, and P. Frossard, "Chebyshev polynomial approximation for distributed signal processing," in *Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011 International Conference on*, June 2011, pp. 1–8.

[17] L. Hubert and P. Arabie, "Comparing partitions," *Journal of Classification*, vol. 2, no. 2-3, pp. 193–218, 1985.

[18] M. Seifi, I. Junier, J. Rouquier, S. Iskrov *et al.*, "Stable community cores in complex networks," *Complex Networks*, pp. 87–98, 2013.