COMPACT ROTATION INVARIANT IMAGE DESCRIPTORS BY SPECTRAL TRIMMING

Maxime Taquet[1,2]*, Laurent Jacques*[2]*, Benoît Macq*[2]*, Sylvain Jaume*[3]

[1]Computational Radiology Laboratory, Children's Hospital, Harvard Medical School, Boston, USA

[2]ICTEAM Institute, ELEN department, Université catholique de Louvain, Belgium

[3]Reid Lab, Department of Neurobiology, Harvard Medical School, Boston, USA

ABSTRACT

Image descriptors are widely used in applications such as object recognition, pattern classification and image registration. The descriptors encode the local visual content of the image to provide a compact, robust and distinctive representation of objects. If images differ in orientation, descriptors must be rotation invariant. This paper introduces a compact rotation invariant descriptor. The approach is based on the representation of the local visual content by a graph. A function living on the graph vertices is evaluated and transformed through spectral trimming. This transform is rotation invariant and reduces the dimensionality of the descriptor. The performance of the introduced descriptor is as good as the SIFT descriptor performance, while being about ten times more compact, as shown by experiments on transmission electron microscope images.

***Index Terms*—** Compact descriptor, invariant features, spectral trimming, Transmission Electron Microscope

1. INTRODUCTION

Descriptors are vectors of parameters describing the local texture or structure of an image. They are extensively used in texture analysis, pattern classification and image registration. These applications are usually dealt with in two separate steps: the detection of key-points and the description of key-point neighborhoods by a vector of parameters. This paper focuses on the second step while a review of key-points detection methods can be found in [1].

Rotation invariance of descriptors is required if scenes may differ in orientation. Three common approaches achieve rotation invariance: histogram-based methods, inner-product methods and dominant orientation detection.

Histogram-based methods sort the values of some property of the key-point neighborhood in a histogram. The resulting histogram is invariant if the property itself is invariant. However, this approach requires a large number of dimensions to achieve good distinctiveness [2].

The inner product between any function and a rotation invariant function is rotation invariant. This property is used in moment-based descriptors [3], descriptors using complex filters [1] and polar Fourier analysis [4]. These methods typically require a large set of invariant functions resulting in high dimensional descriptors and are not robust to small distorsions.

Different methods [5, 6, 2, 7, 8] compute the descriptor with respect to an explicit local *dominant orientation*. SIFT descriptors [5] achieve good performance but are typically high dimensional. To reduce the SIFT dimension, two extensions were proposed: PCA-SIFT [6] and GLOH [1] which both apply principal component analysis (PCA) on the SIFT vectors. The use of PCA requires the prior construction of a database representative of the descriptor subspace and assumes this subspace is euclidean. Furthermore, relying on a single parameter (the dominant orientation) may be misleading if this parameter cannot be robustly defined.

This paper investigates another way to achieve rotation invariance, namely spectral trimming [9]. The key-point neighborhood is represented by an attributed graph. The Graph Fourier Transform (GFT), i.e. the projection of the attribute vector on the graph Laplacian eigenvectors, is then computed. The GFT has two advantages: first, it is rotation invariant, and second, its first coefficients contain most information about the visual content, reducing the descriptors dimensionality without relying on PCA.

The remainder of this paper is organized as follows. Section 2 describes the developed descriptor. Section 3 presents its performances and compare them with SIFT on Transmission Electron Microscopy (TEM) images. Finally, Section 4 concludes the paper and suggests some future research directions.

2. METHODS

The compact descriptor is obtained by describing the key-point intensity neighborhood as a graph and computing the Graph Fourier Transform of a structure function living on its vertices. This section introduces the definition of the graph

MT thanks the Belgian American Educational Foundation (B.A.E.F.) and the Belgian National Science Foundation (F.R.S-FNRS) for their financial support. LJ is funded by the Belgian Science Policy (Return Grant, BELSPO) joined to the Belgian Interuniversity Attraction Pole IAP-VI BCRYPT

from the neighborhood, then the concept of Graph Fourier Transform, and the choice of a structure function. Spectral trimming, bringing everything together in a compact descriptor, is subsequently described. Finally, the numerical complexity is analyzed.

2.1. Graph Definition

Our approach relies on the definition of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ for each key-point neighborhood. Let $\mathcal{I}(\boldsymbol{n})$ be the image and let every pixel and every key-points be referred by their locations on the image grid ($\boldsymbol{c}$ for key-points, $\boldsymbol{n}$ for other pixels). For any key-point $\boldsymbol{c}_k$, the set of vertices $\mathcal{V}_r(\boldsymbol{c}_k)$ is the set of all pixels lying within a radius r from the key-point:

$$\mathcal{V}_r(\boldsymbol{c}_k) = \big\{\boldsymbol{n} \,\big|\, ||\boldsymbol{n} - \boldsymbol{c}_k|| < r\big\}.$$

We note $\#\mathcal{V}_r = N_v$. A four-connectivity is used, i.e. each pixel (vertex in the graph) is connected to its four neighbors. The edge between vertices $\boldsymbol{n}_i$ and $\boldsymbol{n}_j$ is weighted by a gaussian of the difference between intensities, yielding an adjacency matrix $\mathcal{A}$:

$$(\mathcal{A})_{ij} \;=\; \begin{cases} \exp(-\frac{1}{2\sigma^2}\,\|I(\boldsymbol{n}_i) - I(\boldsymbol{n}_j)\|^2), & \text{if } \boldsymbol{n}_i \sim \boldsymbol{n}_j, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The motivation for this choice of adjacency is the convergence towards the true space geometry when the image resolution increases. In particular, the graph Laplacian:

$$\mathcal{L} = \mathbb{I} - \mathcal{D}^{-1/2}\mathcal{A}\mathcal{D}^{-1/2} \in \mathbb{R}^{N_v \times N_v}, \tag{2}$$

with $\mathcal{D}_{ij} = \delta_{ij} \sum_k \mathcal{A}_{ik}$, tends to the continuous Laplacian [10].

2.2. Graph Fourier Transform

The continuous Fourier transform of a function can be considered as the projection of the function onto the eigenfunctions of the Laplacian operator Δ. Indeed, in 2D, $\exp(i\boldsymbol{\omega}^T\boldsymbol{x})$, for any $\boldsymbol{\omega}$, satisfies $\Delta \exp(i\boldsymbol{\omega}^T\boldsymbol{x}) = -||\boldsymbol{\omega}||^2 \exp(i\boldsymbol{\omega}^T\boldsymbol{x})$.

The graph Laplacian eigenvector basis $\mathcal{B}$ constitutes an orthonormal basis of $\mathbb{R}^{N_v}$:

$$\begin{aligned} \mathcal{B} \;&=\; \{\boldsymbol{v}_j \in \mathbb{R}^{N_v} : 1 \le j \le N_v,\ \mathcal{L}\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j\} \\ &\quad \text{with} \quad \lambda_1 = 0,\ \lambda_j \le \lambda_{j+1} \le 2. \end{aligned} \tag{3}$$

The upper-bound on the eigenvalue is due to the Gerschgorin Disk theorem [11]. For a regular distribution of vertices on an infinite plane, $\mathcal{B}$ coincides with the 2-D Fourier basis. Representing this basis $\mathcal{B}$ by the orthogonal matrix $\mathcal{B} = (\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{N_v}) \in \mathbb{R}^{N_v \times N_v}$, the Graph Fourier Transform (GFT) of a vector $\boldsymbol{f} \in \mathbb{R}^{N_v}$ living on $\mathcal{V}$ is naturally defined as

$$\hat{\boldsymbol{f}} = \mathcal{B}^T \boldsymbol{f}, \quad \text{or} \quad \hat{f}_j = \boldsymbol{v}_j^T\, \boldsymbol{f}, \quad \forall 1 \le j \le N_v. \tag{4}$$

(a) (b) (c) (d) (e)

Fig. 1. Key-point neighborhood (a), eigenvector corresponding to $\lambda_1 = 0$ (b), $\lambda_2 = 0.0019$ (c), $\lambda_{10} = 0.011$ (d), $\lambda_{780} = 1.98$ (e). Grayscale is arbitrary, eigenvectors are normalized.

The development of $\mathcal{L}\boldsymbol{v} = \lambda\boldsymbol{v}$ reveals the analogy with the continuous Fourier Transform. Using (2), we get

$$v(i) = \frac{1}{1-\lambda} \sum_{j \sim i} \frac{\mathcal{A}_{ij}}{\sqrt{\mathcal{D}_{ii}\mathcal{D}_{jj}}} v(j), \tag{5}$$

where $j \sim i$ is the condition that j is connected to i. For the lowest eigenvalues, $\lambda \approx 0$, the factor $(1-\lambda)^{-1} \approx (1+\lambda)$. In homogeneous intensity areas ($\mathcal{A}_{ik} \approx 1$, $\mathcal{D}_{jj} \approx \mathcal{D}_{ii} \approx 4$), $v(i)$ tends to be constant and transitions occur at the borders of the region (Fig. 1(b-d)). Projections on eigenvectors with low λ may thereby be considered as a low frequency analysis of the structure function[1]. Symmetrically, for high $\lambda(\approx 2)$, the factor $(1-\lambda)^{-1} \approx -1$ and in homogeneous areas, $v(i)$ is close to the negative mean of the neighboring values (Fig. 1(e)). Projections on these eigenvectors can then be considered as a high frequency analysis of the vector $\boldsymbol{f}$.

Interestingly, the GFT is invariant under any relabeling of the graph vertices. Indeed, given a permutation matrix $\Pi \in \{0,1\}^{N_v \times N_v}$ with only one 1 per row and column and $\Pi^{-1} = \Pi$, if the nodes of $\mathcal{V}$ are permuted accordingly, $\boldsymbol{f} \to \Pi\boldsymbol{f}$, $\Delta \to \Pi\,\Delta\,\Pi^T$ and $\hat{\boldsymbol{f}} \to (\Pi\mathcal{B})^T\Pi\,\boldsymbol{f} = \hat{\boldsymbol{f}}$. After rotation of the image, only the vertex labels are modified. Therefore, the relabeling invariance of the GFT implies its rotation invariance (especially for the first eigenvectors since they are less affected by interpolation errors).

2.3. Structure Function

There exists an infinite choice of structure function $\boldsymbol{f}$. This paper focuses on one particular choice, the vertex degrees:

$$\boldsymbol{f}_{\boldsymbol{c}_k}(\boldsymbol{n}_i) = \mathcal{D}_{ii}$$

Due to equation (1), the degree of vertices take values in $[0, 4]$. For binary images, supposing that $\sigma \to 0$ in equation (1), the vertex degree is truly related to the local structure in the image (Fig. 2). An isolated point has a zero degree; a line edge has a degree of one; the rest of the line and the corner of an object have a degree of two; the edge of an object has a degree of three; and the interior of a object and the crossing of two lines

[1]The notion of frequency here is related to oscillations on the intensity graph. Low frequency is therefore not inconsistent with sharp transitions occurring at the borders of homogenous areas.

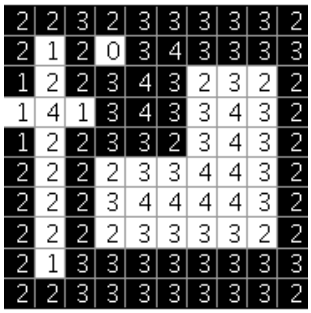

Fig. 2. Nodes degree for a binary image with $\sigma \to 0$ showing the principal recognizable structures.

have a degree of four. For grayscale images, the degrees continuously vary between zero and four and a similar relation with the local structure applies. This function encodes the structure of the object present in the image, hence its name of *structure function*.

2.4. Spectral trimming

Given the structure function $\boldsymbol{f}$ defined in the previous section, its GFT $\hat{\boldsymbol{f}}$ is computed using equation (4). The normalized eigenvectors of the Laplacian are defined up to their sign and so are the coefficients of the GFT. The absolute values of these coefficients are therefore recorded. Since the structure function is presumed to vary smoothly on the neighborhood, the first N_s coefficients ($N_s < N_v$) of its GFT contain most information. The spectral trimming is therefore defined as:

$$\boldsymbol{\varphi}_{\mathrm{Sp}}(\boldsymbol{c}_k) \;=\; \big(|\hat{f}_{\boldsymbol{c}_k}(i)|\big)_{i=1,\ldots,N_s} \;\in\; \mathbb{R}^{N_s}.$$

The operation of trimming the GFT can be interpreted as an intrinsic dimensionality reduction. This vector of N_s parameters constitutes our descriptor. The distance between descriptors is simply the euclidean distance between the vectors φ_{Sp}.

2.5. Complexity

The computational complexity of the descriptor evaluation is split as follows. The time-consuming part of the graph definition is the connectivity estimation in $\mathcal{O}(4N_v)$. The eigenvector decomposition is bounded by $\mathcal{O}(N_v^3)$. However, since the Laplacian is sparse, its first few eigenvectors can be computed efficiently by means of Lanczos algorithms, whose complexity depends on the spectral gap $|\lambda_{k+1} - \lambda_k|$. Finally the spectral trimming has a complexity of $\mathcal{O}(N_s N_v)$. On a 2.8GHz laptop, for $N_v = 793, N_s = 10$, the descriptor construction takes about 0.05s.

3. EXPERIMENTS AND RESULTS

This section assesses the descriptor performance and compares it to SIFT descriptors [5].

Experimental setup: Four 1300×1300 TEM images of a mouse visual cortex with intensities normalized in $[0, 1]$ are used. In each image, 130 key-points are detected. Each image then undergoes 18 rotations uniformly sampled between 0 and 170°. The key-point coordinates are simply rotated along with the images so that the true correspondences are known. This does not bias the comparison results since the same is applied for both the proposed descriptors and SIFT. The descriptors are then computed in every rotated version independently, forming a database of 9360 descriptors containing 520 equivalence classes of 18 descriptors.

For the proposed descriptors, the key-points are detected by a Laplacian of Gaussian (LoG) method [1] with scale parameters varying between $\sigma_{\text{min}} = 2$ and $\sigma_{\text{max}} = 15$. The radius $r = 16$ leads to a number $N_v = 793$ of vertices, and $\sigma = 0.04$ for the adjacency matrix definition.

SIFT descriptors are computed with the parameters suggested in [5]. The key-points detection integrated in SIFT is used rather than LoG, ensuring optimal performance. The detected scales of key-points are also propagated to the rotated images. The dimensionality is a function of the number of spatial and orientation histogram bins.

Performance criterion: The equivalence between two descriptors is tested by comparing the euclidean distance (d) between them to a threshold $\mathcal{T}$ (positive test if $d < \mathcal{T}$ and negative test if $d \geq \mathcal{T}$). Two types of errors may occur: false positives and false negatives. False negatives (positives) occur when the test is negative (positive) although the descriptors are equivalent (distinct).

The number of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are computed for all possible values of $\mathcal{T}$. The Receiver Operating Characteristics (ROC) curve presents the true positives rate ($\text{TPR}(\mathcal{T}) = \text{TP}/(\text{TP} + \text{FN})$) versus the false positives rate ($\text{FPR}(\mathcal{T}) = \text{FP}/(\text{FP}+\text{TN})$). The area under the ROC curve (AUC) equals the probability that the distance between a random pair of equivalent descriptors is lower than the distance between a random pair of distinct descriptors [12]. The AUC is an elegant criterion as it brings down the curve to a single criterion.

Results and Discussion: The curve of the AUC versus the dimensionality shows that our method has better performance with fewer parameters (Fig. 3). A peak of 96.8% is reached for $N_s = 9$. Keeping additional GFT coefficients results in slightly worse results due both to the higher frequency of the corresponding eigenvectors (higher frequencies are subject to significant interpolation errors) and to the instability of the Lanczos algorithm for higher eigenvalues. The plateau reached by the SIFT performance may be related to errors in the dominant orientation detection which is not reduced by an increased dimensionality.

The observation of a key-point neighborhood and its nearest neighbors (in terms of the distance between their descriptors) among all key-points detected in a rotated version of

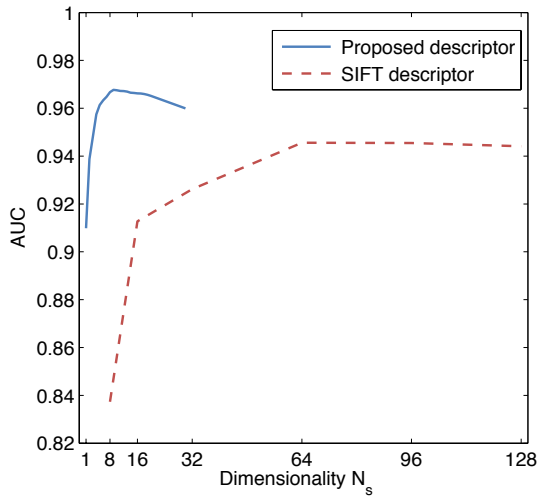

Fig. 3. Evolution of the AUC with the descriptor dimensionality for the proposed descriptor and for SIFT descriptors.

the image, offers a better insight of the descriptors behavior (Fig. 4). We notice that neighborhoods that have the same structure contents are close together.

4. CONCLUSION AND FUTURE WORK

In this paper, spectral trimming has been used to produce compact rotation invariant image descriptors. On top of ensuring the rotation invariance, spectral trimming offers an intrinsic dimensionality reduction based on the low frequency content of the structure function. As a result, better rotation invariance is achieved with a dimensionality about ten times lower than SIFT descriptors. Unlike other compact descriptors, spectral trimming does not require a prior PCA and makes no assumption on the descriptor subspace.

In a future work, we want to use our descriptors to register 2D TEM images in order to reconstruct a volumetric image. The invariance of the descriptors under anatomical changes should be investigated. The descriptor could also be extended to be invariant under a larger class of transformations. Scale invariance could be obtained by adaptively fixing the radius r based on the scale detected by the LoG detector. Similarly, invariance to smooth contrast changes could be obtained by adaptively fixing the factor σ based on the local intensities.

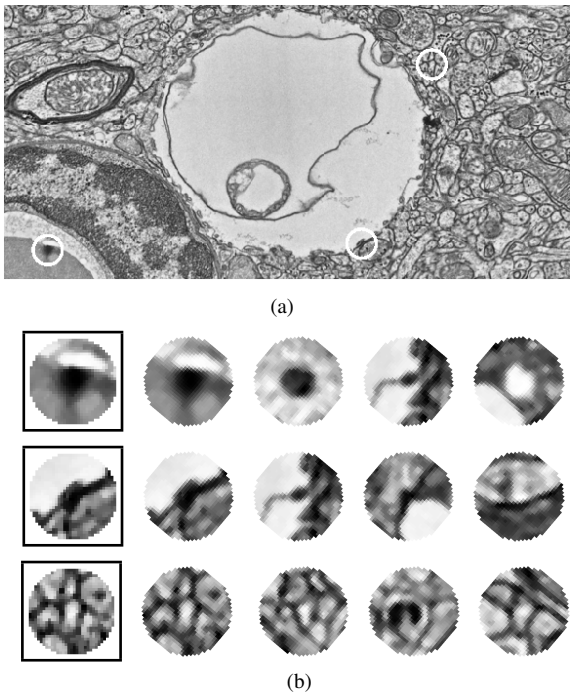

Fig. 4. (a) Part of a TEM image with three detected key-points (circled on the image). (b) Left to right: the three neighborhoods detected in (a) and the four nearest neighbors among the 130 candidates in a 40° rotated version of the image (they were rotated backward for the sake of visualization).

5. REFERENCES

[1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L.V. Gool, "A comparison of affine region detectors," *IJCV*, vol. 65, no. 1, pp. 43–72, 2005.

[2] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE PAMI*, pp. 1615–1630, 2005.

[3] J. Heikkilä, "Pattern matching with affine moment descriptors," *Pattern recognition*, vol. 37, no. 9, pp. 1825–1834, 2004.

[4] Q. Wang, O. Ronneberger, and H. Burkhardt, "Rotational invariance based on Fourier analysis in polar and spherical coordinates," *IEEE PAMI*, vol. 31, no. 9, pp. 1715–1722, 2009.

[5] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91–110, 2004.

[6] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," *CVPR*, 2004.

[7] Z. Chen and S.K. Sun, "A zernike moment phase-based descriptor for local image representation and matching," *IEEE TIP*, vol. 19, no. 1, pp. 205–219, 2009.

[8] M. Mellor, B.W. Hong, and M. Brady, "Locally rotation, contrast, and scale invariant descriptors for texture analysis," *IEEE PAMI*, pp. 52–61, 2007.

[9] M. Taquet, L. Jacques, C. De Vleeschouwer, and B. Macq, "Invariant Spectral Hashing of Image Saliency Graph," *Proc. Computational Topology in Image Context*, pp. 105–112, 2010.

[10] L. Jacques, "Convergence Rate of the Symmetrically Normalized Graph Laplacian," Tech. Rep., ICTEAM/UCL, 2011, arxiv:1101.1428.

[11] F.R.K. Chung, "Spectral Graph Theory," *Regional Conference Series in Mathematics, American Mathematical Society*, vol. 92, pp. 1–212, 1997.

[12] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.