

COMBINING SHAPE DESCRIPTORS AND COMPONENT-TREE FOR RECOGNITION OF ANCIENT GRAPHICAL DROP CAPS

Benoît Naegel and Laurent Wendling

LORIA UMR 7503, Nancy University, 54506 Vandœuvre-lès-Nancy, France

Keywords: Component-tree, Segmentation, Classification, Mathematical morphology, Ancient graphical drop caps.

Abstract: The component-tree structure allows to analyse the connected components of the threshold sets of an image by means of various criteria. In this paper we propose to extend the component-tree structure by associating robust shape-descriptors to its nodes. This allows an efficient shape based classification of the image connected components. Based on this strategy, an original and generic methodology for object recognition is presented. This methodology has been applied to segment and recognize ancient graphical drop caps.

1 INTRODUCTION

Many pattern recognition tasks rely on thresholding techniques, which can be very efficient to perform two class classification between object and background pixels (Weszka, 1978; Fu and Mui, 1981; Sahoo et al., 1988; Trier and Jain, 1995; Trier and Taxt, 1995a; Sezgin and Sankur, 2004). However these methods all rely on the choice of a threshold, either global or local. Therefore, to avoid the threshold computation, one strategy could consist to consider the connected-components of all the image threshold sets. These connected-components can be organized in a tree called “component-tree”. Component-trees (also called max-tree (Salembier et al., 1998), dendrone (Chen et al., 2000), confinement-tree (Mattes and Demongeot, 2000)) have been involved in many image processing tasks, such as image simplification (Salembier et al., 1998), object detection (Jones, 1999; Naegel et al., 2007), image retrieval (Mosorov, 2005), caption text detection (León et al., 2005). The component-tree has been used in mathematical morphology to perform attribute filtering (Breen and Jones, 1996), by storing an attribute into each node of the tree. The notion of shape-based attributes has been considered only recently (Urbach and Wilkinson, 2002; Urbach, 2005; Urbach, 2007) and this concept has been applied to filament extraction in MR angiograms (Urbach and Wilkinson, 2002) and classification of diatoms (Urbach, 2007). For both segmentation and detection purposes, shape-descriptors can be used to perform efficient shape-based classification of

the nodes. In this paper we expose a method allowing to perform both object segmentation and recognition based on the component-tree structure. Such an approach has been experimented to perform ancient drop cap recognition.

2 COMPONENT-TREE FOR OBJECT RECOGNITION

2.1 Component-tree Definition

An attributed component-tree is defined as a triple $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \delta)$, where \mathcal{V} represents the set of nodes, \mathcal{E} is the set of edges and δ a function that assigns to each node $u \in \mathcal{V}$ a set of attributes $\delta(u)$. A grey-level image is defined as a function $F : E \rightarrow T$, where E is the domain of definition ($E \subseteq \mathbb{Z}^2$) and $T \subseteq \mathbb{Z}$ is the set of values. As we consider only discrete images having discrete values, we set $T = [0, M]$. The i -th connected component of the superior threshold sets $X_i(F) = \{p \in E \mid F(p) \geq t\}$ is denoted $C_i^t(F)$ (in the sequel we write C_i for designing such a component). The set of all connected components of all superior threshold sets is denoted $\mathcal{CC}(F) = \{C_i^t(F) \mid t \in T, i \in I\}$ (this set is denoted \mathcal{CC} in the sequel). Inferior threshold sets $X^u(F) = \{p \in E \mid F(p) \leq u\}$ can be considered as well, leading to the dual component-tree representation. The component-tree is constructed as follows. A map $m : \mathcal{V} \mapsto \mathcal{CC}$ is defined between the set of vertices and the set of components. For each $C_i \in \mathcal{CC}$,

a node $u \in \mathcal{V}$ associated to this component is created. An edge is created between each pair of vertices $(u, v) \in \mathcal{V} \times \mathcal{V}$ if and only if:

- (i) $u = m^{-1}(C_t), v = m^{-1}(C_{t+k})(t, k \geq 0)$;
- (ii) $C_{t+k} \subsetneq C_t$;
- (iii) $\forall l \geq 0, C_{t+l} \subsetneq C_t \Rightarrow C_{t+l} \subsetneq C_{t+k}$,

and u is the father of v . The root of the tree is denoted by the node $r = m^{-1}(C_0)$. An example of component-tree constructed from a 1D image is provided in Figure 1.

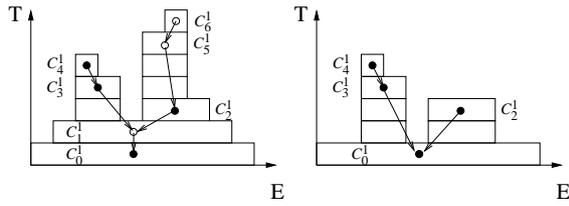


Figure 1: a) Component-tree of a 1D image. Empty circles denote nodes that do not meet a criterion, b) Tree pruning. In the reconstructed image (using *direct* decision), all irrelevant components C_t have been removed.

2.2 Node Classification

Given a criterion D that accepts or rejects a node depending on its attribute values, an attribute filter ϕ can be defined by acting separately on the component-tree nodes. Pruning the tree according to this classification decision and constructing the image obtained from the pruned tree allows to remove all the irrelevant connected components of the threshold sets of an image. This is the principle of attribute filters (Breen and Jones, 1996; Jones, 1999). Pruning the component-tree with respect to non-increasing criterion can lead to remove nodes which have ancestors or descendants that are not removed, which is problematic to reconstruct an image. Therefore, various strategies have been defined to reconstruct an image from a tree pruning based on non-increasing criterion: *min*, *max*, *direct*, *subtractive* (Salembier et al., 1998; Urbach, 2007). In an object detection purpose, image reconstruction can be based on the *direct* decision (Salembier et al., 1998), which is equivalent to keep only the relevant connected components of the image (see Figure 1): $\phi(F)(x) = \max\{t \mid x \in C_t, D(m(C_t)) = \text{true}\}$.

2.3 Attributes

A set of attributes $\delta(u)$ is associated to each node $u \in \mathcal{V}$. For efficiency reasons, these attributes should be computed incrementally: for each node, the computation result of child attributes is used to compute the current ones.

2.3.1 Scalar Attributes

Scalar attributes of various kinds can be associated to nodes. Many attributes have been used in the literature such as component area, length of the shortest path to a leaf node, cumulated sum of area of all descendant nodes. Pruning the tree according to criteria based on these attributes led respectively to the area filter (Vincent, 1992; Vincent, 1993), the h-min filter (Grimaud, 1992; Soille, 2003), and volumic filter (Vachier, 1998), well known in mathematical morphology. Combination of scalar attributes can also be used, leading to vectorial attributes (Urbach, 2005; Urbach, 2007; Naegel et al., 2007). However these kinds of quantitative attributes are not descriptive enough to perform a robust shape-based classification of nodes.

2.3.2 Shape Descriptors

To use the component-tree for pattern recognition, shape descriptors can be associated to nodes. Indeed, shape descriptors allow compact description of a shape, while being robust against noise and pose variations. There is a large choice of shape descriptors available in the literature (see (Zhang and Lu, 2004) for a survey). In this paper, we compare the performance of two descriptors, chosen for their robustness and their performance in content-based image retrieval systems: the R-transform and the generic Fourier descriptor (GFD).

R-Transform. The R-transform (Tabbone et al., 2006) is a shape descriptor based on the Radon transform. The Radon transform of an image $f(x, y)$ is defined by:

$$T_{Rf}(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - \rho) dx dy,$$

where δ is the Dirac delta-function ($\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ elsewhere), $\theta \in [0, \pi[$ and $\rho \in]-\infty, \infty[$. The R-transform is defined by:

$$\mathcal{R}_f(\theta) = \int_{-\infty}^{\infty} T_{Rf}^2(\rho, \theta) d\rho$$

The R-transform is invariant under translation and scaling if the transform is normalized by a scaling factor (the area of the R-transform). However the R-transform is not invariant under rotation (a rotation of the shape implies a translation of the transform).

Generic Fourier Descriptor. Generic Fourier descriptors (GFD) (Zhang and Lu, 2002) belong to standard MPEG-7. They are defined from polar discrete

transform. It is similar to Fourier transform, but considering the polar image in a polar space as a rectangular image in a Cartesian space:

$$PF(\rho, \phi) = \sum_x \sum_y f(x, y) \cdot e^{[2j\pi \left(\frac{r(x,y)}{R} \rho + v(x,y)\phi \right)]}$$

The *GFD* is defined as follows:

$$GFD = \left\{ \frac{|PF(0,0)|}{M_{11}}, \frac{|PF(0,1)|}{|PF(0,0)|}, \dots, \frac{|PF(m,n)|}{|PF(0,0)|} \right\}$$

GFD is invariant to rotation and translation by definition. Invariance to scale factor is ensured by dividing the first term by area value M_{11} and the others by $|PF(0,0)|$.

Incremental Computation. Computing shape descriptors for all the threshold components leads to a high computational burden. In order to compute efficiently GFD or R-transform for each node, it is desirable to find an incremental scheme allowing to use the results of the child nodes to compute the result of the current node. This can be done by using accumulator arrays. The computation of the GFD was based on an incremental scheme, by taking as origin the image barycenter for all the nodes.

3 RECOGNITION OF ANCIENT GRAPHICAL DROP CAPS

In recent years, large digitalization campaigns of ancient printed document collections led to recurrent problems such as data archiving, classification and efficient retrieval. Due to the handwritten nature of ancient books, recognition of ancient characters cannot be easily achieved by using modern optical character recognition (OCR) systems. Ancient graphical drop caps segmentation has been addressed by Uttama (Uttama et al., 2005), where the image is separated into its textural and homogeneous parts. However the problem of drop cap recognition has not been addressed due to the difficulty of extracting the drop cap's uppercase.

3.1 Data

The Madonne database OLDB (Ornamentals Letters DataBase) consists of graphical decorative initials extracted from archival documents¹. This corpus is part

¹We would like to thank the "Centre d'études Supérieures de la Renaissance" for the permission to use their archival documents.

of a study performed by French research teams² on historical documents. The database is composed of more than 6000 drop caps. To conduct our experiments, we have selected from this database a subset of 200 images representing 20 different uppercases. The test set has been constructed to be representative of the various styles of drop caps (see examples on Figure 2). Drop cap images are grey-scale, composed of the letter (uppercase) part and texture part. Size of images is between 150×150 and 750×750 . Due to the digitalization process, the letter is approximately centered and vertical. Drop cap images are noisy and contain artifacts such as superimposed text, coming from following book pages.



Figure 2: Drop cap samples of a subset of the OLDB (Ornamentals Letters DataBase) database.

3.2 Image Processing

Using the methodology exposed in Section 2 for drop cap recognition requires to make the assumption that at least one connected component C_i of the superior threshold sets $X_i(F)$ corresponds to the uppercase part. By making this assumption, the letter part of the drop cap must appear white on dark background. This assumption does not always hold, since drop caps uppercases can also appear black on a white background (see Figure 2). It is hence necessary to consider not only the connected components of the superior threshold sets, but also those of the inferior threshold sets. To this aim, the dual component-tree (the component-tree of the inferior level sets) is computed for each image. It is obtained by computing the component-tree of the negative image. In the test set, many drop caps contain a distinct component corresponding to the letter part. However in some cases the letter part is connected to other image parts (texture or background). In order to ensure that at least one component corresponds to the letter part we process each image using morphological openings (Soille, 2003), to remove thin connections between components. Due to the stack property of morphological operators, applying a morphological grey-scale opening is equivalent to applying a binary opening on all of its superior threshold sets. Therefore, for each drop cap image F , we compute the set:

$$S(F) = \{F, F^c, \{\gamma_{B_i}(F)\}_{1 \leq i \leq r_{max}}, \{\gamma_{B_i}(F^c)\}_{1 \leq i \leq r_{max}}\},$$

²see <http://l3iexp.univ-lr.fr/madonne/>

where F^c is the image complement (negative), γ denotes the morphological opening, $B_i \in \mathcal{P}(E)$ is the disk structuring element of radius i , and r_{max} is the maximal value of the radius. For each image $F_i \in S$ is computed the corresponding component-tree G_i .

From an algorithmic point of view, Salembier's algorithm (Salembier et al., 1998) was used to compute efficiently the component-tree, using a recursive scheme³.

3.3 Training Set

A set of letters has been segmented using interactive method based on the choice of the most relevant connected component of the threshold sets, using area and compacity criteria to prune the component-tree.

The GFD signatures associated to the component-tree nodes have been computed using an efficient incremental method (Section 2). As a consequence, the GFD signatures are not translation invariant, since the same origin has been taken for all the components. Although the uppercases are approximately centered, slight translation variations can occur between letters of similar classes. This can be addressed by adding in the training set translated versions of segmented letters along the x-axis.

We chosed a cluster head for each class and computed two translated images by a vector $v = (v_x, 0)$, with $v_x = \{\pm 0.1 dx\}$, dx denoting the image width. Each class of the training set is then composed of a segmented image and its two translations.

3.4 Recognition Process

For each tree node u is retrieved the least Euclidian distance between the node attributes and the training set samples attributes. From all these distances, the k least distances are retained, as well as the k corresponding classes. The recognition is performed by considering the most represented class from these k classes (here $k = 30$).

4 RESULTS

4.1 Validation Protocol

We have compared the following methods. The three first ones rely on the component-tree computation; the

³According to (Najman and Couprie, 2006), Salembier's algorithm is quadratic in the worst case; however it is generally twice as fast as Najman's one in practical cases when the image range is $[0 \dots 255]$.

fourth is based on global grey-scale GFD computation. The last one is based on a binarization strategy.

1. **CT+scalar.** Component-tree based recognition using scalar attributes. In this method, we consider for each node $u \in \mathcal{V}$ the set of scalar attributes $\delta(u) = (a_1, a_2)$ with:

- $a_1 = \text{area}/(\text{image size})$ the normalized area of the component with respect to the image size;
- $a_2 = 4\pi * \text{area}/\text{perimeter}^2$ an attribute related to the compacity of the component.

2. **CT+GFD.** Component-tree based recognition using GFD. The GFD has been computed using $m = 5$ radial and $n = 5$ angular frequencies.

3. **CT+RT.** Component-tree based recognition using R-Transform. The R-Transform has been computed using $r = 10$ radial and $t = 10$ angular samples.

For the GFD and the R-Transform, a performance evaluation of the shape descriptors following different discretization parameters has been performed. In both cases, increasing the parameters beyond the chosen ones did not improve the results while increasing the processing time. In the latter methods based on the component-tree, the maximal radius for the morphological openings has been set to $r_{max} = 3$.

4. **Grey-scale GFD.** Classification based on grey-scale GFD computation. A training set composed of one head drop cap for each class has been constructed. Each tested image was classified according to the closest training set GFD.

5. **Entropy+GFD.** Binarization method based on entropy. A powerful binarization method which performs a fuzzy partition on a bidimensional histogram of the image (Cheng and Chen, 1999) has been experimented. The partition criteria are based on the optimization of the fuzzy entropy. This approach is threshold free and it has been shown that the results obtained are better than for the 2D non fuzzy approach. The image is classified according to the GFD of the remaining connected components, using the same method (training set and classification strategy) than CT+GFD.

4.2 Results on 8 Classes

In a first experiment, we tested the methods on a subset of 80 drop caps divided in 8 classes. The selected classes correspond to the most frequent letters in the whole database. Table 1 shows a comparison of the segmentation and classification results between methods.

Table 1: Comparison of segmentation, recognition rates and computational time (in s) of methods on 80 drop caps from 8 classes.

| Method | Segmentation | Recognition | Time |
|-----------|--------------|--------------|------|
| CT+scalar | 7.5% | 12.5% | 52 |
| CT+RT | 5% | 18.8% | 575 |
| CT+GFD | 62.5% | 76.3% | 332 |
| ent.+GFD | 38.7% | 47.5% | 534 |
| gr. GFD | - | 18.8% | 40 |

The method based on the component-tree and GFD descriptor outperformed the others, with a recognition rate of 76.3%. This result is promising, since the method has not been specifically tuned for the drop cap recognition. Component-tree methods based on other attributes (scalar or R-transform) performed poorly on the considered dataset. The method based on entropy binarization gave interesting results; however in 25% of the images, the method chose a wrong threshold, leading to nearly black or white images. Moreover this method is slower than CT+GFD, due to the computation of the optimal threshold. Grey-scale GFD gave poor results, since the drop caps show many texture and contrast variations. The confusion matrix obtained with the CT+GFD method is given in Table 2.

Table 2: Confusion matrix of the classification results using the CT+GFD method on 8 classes.

| Uppercase | A | C | E | I | L | P | Q | S |
|-----------|----------|----------|----------|----------|----------|----------|-----------|----------|
| A | 8 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| C | 0 | 8 | 0 | 0 | 0 | 0 | 2 | 0 |
| E | 0 | 0 | 6 | 1 | 1 | 0 | 1 | 1 |
| I | 0 | 0 | 0 | 7 | 1 | 0 | 0 | 2 |
| L | 0 | 1 | 0 | 1 | 7 | 0 | 0 | 1 |
| P | 0 | 0 | 2 | 0 | 2 | 6 | 0 | 0 |
| Q | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 |
| S | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 9 |

In the tree based methods, morphological opening allowed to disconnect the uppercase from texture parts as illustrated on figure 3, therefore increasing the recognition rate.



Figure 3: a) Original drop cap image, b) Segmentation of (a) (letter recognized as "S"), c) Opened image (the uppercase is disconnected from the background), d) Segmentation of (c) (letter properly recognized).

4.3 Scalability of the Approach

The number of classes has been gradually increased, by decreasing order of their frequencies in the whole database. The Figure 4 summarizes the results obtained with the different methods and shows the good behaviour of the proposed approach.

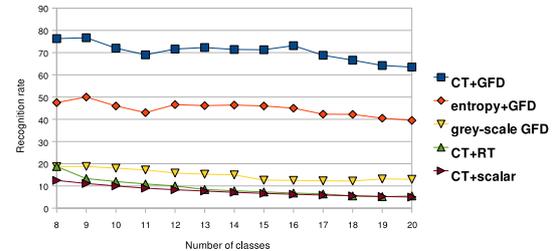


Figure 4: Recognition percentages of the methods with respect to the number of considered classes.

4.4 Discussion

The method based on entropy binarization and GFD classification is less robust than the method CT+GFD, due to the inaccurate image binarization in some cases. These results demonstrate that method based on the component-tree structure extended with robust shape-descriptors allows to analyse all the threshold components, therefore avoiding to chose a specific binarization threshold. Hence, the method CT+GFD reached the best results, achieving a recognition rate of 63.5% on 20 classes.

5 CONCLUSIONS

In this paper we have proposed to combine the component-tree representation of an image with robust shape-descriptors, in order to perform object recognition. Based on this strategy, we have exposed a methodology for ancient graphical drop cap recognition. The best recognition rate was 76.3% on 8 classes and 63.5% on 20 classes. These results, although perfectibles, are interesting since the method has not been specifically tuned for this application. Moreover, the computation of the attributed component-tree could be done offline, leading to efficient recognition tasks.

In future works we will investigate the classification strategies offered by the combination of component-tree and shape descriptors. More specifically, we plan to use more deeply the information carried by the component-tree structure to perform image recognition and spotting.

REFERENCES

- Breen, E. and Jones, R. (1996). Attribute openings, thinnings, and granulometries. *CVIU*, 64(3):377–389.
- Chen, L., Berry, M., and Hargrove, W. (2000). Using dendronal signatures for feature extraction and retrieval. *International Journal of Imaging Systems and Technology*, 11(4):243–253.
- Cheng, H.-D. and Chen, Y.-H. (1999). Fuzzy partition of two-dimensional histogram and its application to thresholding. *Pattern Recognition*, 32(5):825–843.
- Fu, K. S. and Mui, J. K. (1981). A Survey on Image Segmentation. *Pattern Recognition*, 13:3–16.
- Grimaud, M. (1992). New measure of contrast: dynamics. In Gader, P., Dougherty, E., and Serra, J., ed., *Image Algebra and Morphological Image Processing III*, vol. SPIE-1769, pages 292–305. SPIE.
- Jones, R. (1999). Connected filtering and segmentation using component trees. *CVIU*, 75(3):215–228.
- León, M., Mallo, S., and Gasull, A. (2005). A tree structured-based caption text detection approach. In *Fifth IASTED International Conference Visualisation, Imaging and Image Processing*, pages 220–225.
- Mattes, J. and Demongeot, J. (2000). Efficient algorithms to implement the confinement tree. In Borgefors, G., Nyström, I., and di Baja, G. S., ed., *DGCI'00*, vol. 1953 of *LNCIS*, pages 392–405. Springer.
- Mosorov, V. (2005). A main stem concept for image matching. *Pattern Recognition Letters*, 26:1105–1117.
- Naegel, B., Passat, N., Boch, N., and Kocher, M. (2007). Segmentation using vector-attributes filters: methodology and application to dermatological imaging. In Bannon, G., Barrera, J., and Braga-Neto, U., ed., *ISMM 2007, Brazil*, vol. 1, pages 239–250. INPE.
- Najman, L. and Couprie, M. (2006). Building the component tree in quasi-linear time. *IEEE Trans. on Image Processing*, 15(11):3531–3539.
- Sahoo, P. K., Soltani, S., Wong, A. K. C., and Chen, Y. C. (1988). A survey of thresholding techniques. *CVGIP*, 41(2):233–260.
- Salembier, P., Oliveras, A., and Garrido, L. (1998). Anti-extensive connected operators for image and sequence processing. *IEEE Trans. on Image Processing*, 7(4):555–570.
- Sezgin, M. and Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):146–165.
- Soille, P. (2003). *Morphological Image Analysis: Principles and Applications*. Springer-Verlag Berlin Heidelberg, 2nd edition.
- Tabbone, S., Wendling, L., and Salmon, J.-P. (2006). A new shape descriptor defined on the radon transform. *CVIU*, 102:42–51.
- Trier, Ø. and Taxt, T. (1995a). Evaluation of Binarization Methods for Document Images. *IEEE Transactions on PAMI*, 17(3):312–315.
- Trier, Ø. D. and Jain, A. K. (1995). Goal-Directed Evaluation of Binarization Methods. *IEEE Trans. on PAMI*, 17(12):1191–1201.
- Urbach, E. (2005). Vector attribute filters. In *ISMM'05 - International Symposium on Mathematical Morphology*, vol. 30 of *Computational Imaging and Vision*, pages 95–104. Springer SBM.
- Urbach, E. (2007). Connected shape-size pattern spectra for rotation and scale-invariant classification of gray-scale images. *IEEE Trans. on PAMI*, 29(2):272–285.
- Urbach, E. and Wilkinson, M. (2002). Shape-only granulometries and gray-scale shape filters. In *ISMM'02*, pages 305–314. CSIRO Publishing.
- Uttama, S., Ogier, J., and Loonis, P. (2005). Top-down segmentation of ancient graphical drop caps: Lettrines. In *Proceedings of the 6th GREC*, pages 87–96.
- Vachier, C. (1998). Utilisation d'un critère volumique pour le filtrage d'image. In *RFIA '98*, pages 307–315.
- Van Droogenbroeck, M. and Talbot, H. (1996). Fast computation of morphological operations with arbitrary structuring elements. *Pattern Recognition Letters*, 17(14):1451–1460.
- Vincent, L. (1992). Morphological Area Openings and Closings for Grey-Scale Images. In *Shape in Picture: Nato Workshop*, pages 197–208. Springer.
- Vincent, L. (1993). Grayscale area openings and closings, their efficient implementations and applications. In *EURASIP Workshop on Mathematical Morphology and its Applications to Signal Processing*, pages 22–27.
- Weszka, J. S. (1978). A survey of threshold selection techniques. *Computer Graphics and Image Processing*, 7:259–265.
- Zhang, D. and Lu, G. (2002). Shape-based image retrieval using generic Fourier descriptor. *Signal Processing: Image Communication*, 17:825–848.
- Zhang, D. and Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, 37(1):1–19.