Near-duplicate Manga Image Detection with Binary Template Selection and Matching

Mathieu Delalandre[1]
[1]University of Tours
Tours city, France
mathieu.delalandre@univ-tours.fr

The Anh Pham[2]
[2]Hong Duc Unversity
Thanh Hoa city, Vietnam
phamtheanh@hdu.edu.vn

Motoi Iwata[3]
[3]Osaka Prefecture University
Osaka city, Japan
iwata@cs.osakafu-u.ac.jp

***Abstract*—Near-duplicate detection aims to detect images that slightly differ in content. This problem has been addressed for natural and document images in the literature. We are interested here in a particular application dealing with Manga / Comics images. In recent years, there has been a noticeable shift of attention on this problem due to copyright protection and piracy. Piracy is interested to getting copies of Manga / Comics. The usual process is to purchase official versions in a print / digital form, to capture and made them available on the web. The problem of copyright protection is completely related to duplicate image detection. It can can be formulated as the comparison of pages coming from legal copies to pages coming from illegal copies. The major issues here are related to noise coming from the capture process, image recompression or postprocessing. The Manga / Comics images are mixture of natural and document images, and the standard duplicate detection methods are not fully adapted to process them. We propose in this paper to investigate template selection and matching in the binary domain to deal with this problem. Our approach deals with the no-deformable and no-invariant matching case. The first experiments report that our method is robust to noise, scalable, supports skewing and low resolution and is time efficient.**

Keywords*—*near-duplicate image detection, template matching and selection, auto-correlation features, binary similarity measures

I. Introduction

The near-duplicate image detection methods aim to detect the duplicates of a query image. Images that are near-duplicate slightly differ in content. The differences can result from the digitalization process, streaming capture, cropping operations, resampling or recompression processes, post-processing, etc. Several researches have been conducted for addressing near-duplicate image detection problem, both for natural and document images [1], [2]. The selection of a suitable method depends on the kind of application use-case which is to be solved. When the problem is less concerned with geometric invariance (e.g., image rotation, shearing, etc.) or partial detection (e.g., finding a crop, dealing with occlusions, etc.), it is referred as an exact duplicate detection in the litterature [1].

Several global descriptors have been proposed for exact duplicate detection of natural images such as GIST, the Haar wavelets, the PCA-transformed histogram or radon-based geometric invariant features [1], [3]. They outperforms local approaches in most of the cases. When dealing with DIA, a combination of global and local features has been investigated in [2] and tested against the literature methods.

In this work we are interested in a particular exact detection application dealing with Manga / Comics images (see Fig. 1). The detection of Manga / Comics duplicate is a core topic for copyright protection and piracy. Piracy is interested to getting copies of Manga / Comics. The usual process is to purchase official versions in a print / digital form and to capture them. The camera-ready versions of illegal copies are then put online on web portals. The detection methods support comparison of pages coming from legal copies to pages coming from illegal copies on the web. The major issues are related to noise coming from the capture process, image recompression or postprocessing. In the last recent years, different contributions have been proposed on this problem for images and/or books copy detection [4], [5], [6].

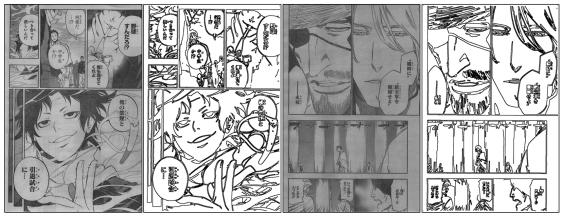

Figure 1. Manga/Comics images with their line drawing content

The exact duplicate detection methods of the CV and DIA fields are not fully adapted to process Comics / Manga images. These are a mixture of natural and document images, they can be mainly characterized as a line drawing content Fig. 1. We propose in this work to investigate template matching and selection in the binary domain to deal with this problem. We exploit the template matching to process the graphic layer of Manga / Comics images, obtained with edge detection, to detect the exact duplicates. Our approach deals with the no-deformable and no-invariant matching case. The first experiments report that our method is robust to noise, scalable, supports skewing and low resolution and is time efficient.

In the rest of the paper we will introduce in section II the problem statement and state-of-the-art about template selection. Section III will provide a description of our approach and section IV will report experiments and results. Conclusions and perspectives will be given in section V.

II. Problem statement and state-of-the-art

Template matching is a well known image processing methodology. Finding a given template in an image is typically

performed by scanning the image and evaluating the similarity between the template and an area. The matching process chooses the template position that maximizes (or minimizes) a given similarity measure between the template and the image. Typical measures are the L_p-norm, the Normalized Cross Correlation (NCC) or the binary similarity measures. Other measures make the matching robust to rotation, affine transformations, occlusions, or illumination variations.

The performance of the template matching process is dependent of the used template. The selection of templates is usually performed manually which is often time consuming and error prone. It is desirable to make the process automatic and this is referred as the template selection problem in the literature [7], [8], [9], [10], [11], [12].

Template selection may be posed as follows. Given an image (or a sub-image) I of size $M \times N$ and a template $X_k \in I$ of size $s \times t$, we have $\complement = (M-s) \times (N-t)$ templates $\in I$. The template selection process aims to identify the template X_k, with $X_1, \ldots, X_k, \ldots, X_\complement$, that best characterizes the image in terms of robustness, scalability and location accuracy.

The standard approach for template selection is to characterize the uniqueness of a template with the crosscorrelation map resulting of the template matching process, and to look for the peak response [9], [11], [12]. This evaluates the crosscorrelation between the template and its duplicate. A strong peak response in the correlation coefficients at the template's position is evidence that the template is unique. The peak response can be characterized through the maximum value of the peak, i.e. the local maxima [9], [12]. Alternative to local maxima is to minimize a noise measure such as the Signal-to-Noise-Ratio (SNR) [11].

The local maxima and SNR require reference images (i.e. a trained dataset) to be computed, and such images cannot be still obtained. This restricts their application to real-life use-cases. To avoid this limitation, no-reference methods can be used. They include shape analysis [7], [8], [10] and autocorrelation features [12]. These methods must be correlated with the local maxima and the SNR measures, as they constitute the optimal characterization of the uniqueness of templates.

Shape analysis can be used to determine the goodness of templates [7], [8], [10]. Features are extracted without the need correlation operations or reference images, they are straightforward to compute. They include entropy [10], the maximum Intensity Variation Number (strongly related to entropy) [7] or the coherence measure that indicates how well the gradients are pointing in the same direction [8].

Shape analysis methods have a little meaning for template selection, as they operate in a different domain compared to the matching process. An alternative is to employ autocorrelation features (i.e. features computed from the crosscorrelation of a template with itself). Autocorrelation features for template selection have been little investigated in the literature. In [12], peak response is characterized by inspecting the shape of the peak. We would want the shape of peak as sharp as possible so that the position of the detected template will be accurate. The autocorrelation features constitute a meaningful estimator for matching. This is coming from the general observation that the matching score $S(X,Y)$ between a template X and a duplicate image Y is a combination of the autocorrelation $S(X,X)$ and noise $\eta(X,Y)$ functions Eq. (1). That is, properties with $S(X,X)$ could drive a minimization of $\eta(X,Y)$. In addition, we will highlight in this paper how the autocorrelation features can support large time optimization for matching. The next section will detail our approach.

$$S(X,Y) = S(X,X) + \eta(X,Y) \tag{1}$$

III. Binary template selection and matching with autocorrelation features

In this section we present our approach for binary template selection and matching using autocorrelation features. Our overall process is introduced in Fig. 2 and detailed in next sections (A.) to (F.). A sampling process extracts $\mathcal{C} \ll \complement$ templates from an image I, by restricting their overlapping (e.g., not more than 90% of overlapping). This rule avoids to process close templates and saves computation time, as a neighboring at some pixels results in a strong template similarity. These templates are then processed to get autocorrelation maps (section A). As we process our images in the binary domain for no-deformable and no-invariant matching, we have considered the binary similarity measures for the autocorrelation and matching processes [13]. Our autocorrelation features are extracted at two levels to characterize the robustness to matching (sections B, C) and the pruning abilities (section D) of templates. Both used input parameters (τ, ω) modeling the expected upper-bound noise of the application use-case. We have reformulated the matching process (section E) to embed the autocorrelation features for time optimization. The final process for selection is then detailed in section F.

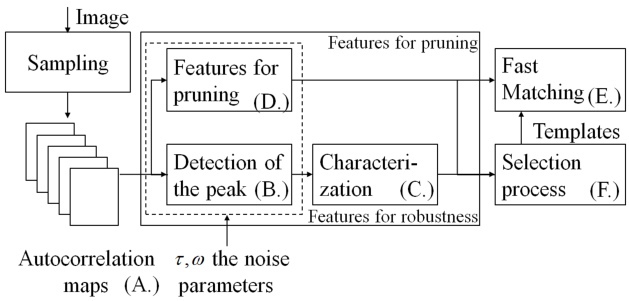

Figure 2. The template selection process

A. The autocorrelation map

Considering a template X of size $s \times t$ (height and width), the autocorrelation map Fig. 3 in a local neighborhood of the template X is a matrix W of size $(2s+1) \times (2t+1)$, where $i_c = s+1, j_c = t+1$ is the center of the map and the location of the peak. This autocorrelation map is computed from a region of interest in I of size $(3s+1) \times (3t+1)$. In this matrix, all elements $W_{i,j}$ provide the similarity measure $S(X, X_{i,j})$ between the template X and the shifted template $X_{i,j}$ using an offset $\Delta_i = i - i_c, \Delta_j = j - j_c$. At $\Delta_i = \Delta_j = 0$ we have then the center of the map i_c, j_c where $S(X, X_{i,j}) = U_B$ is the upper bound of the considered similarity measure.

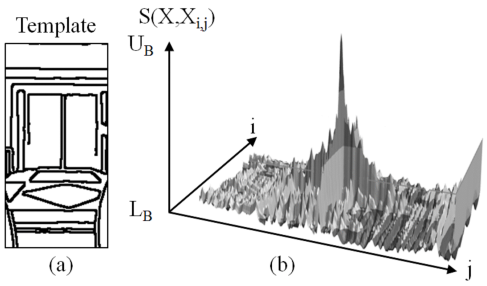

Figure 3. A template (a) and the corresponding autocorrelation map (b)

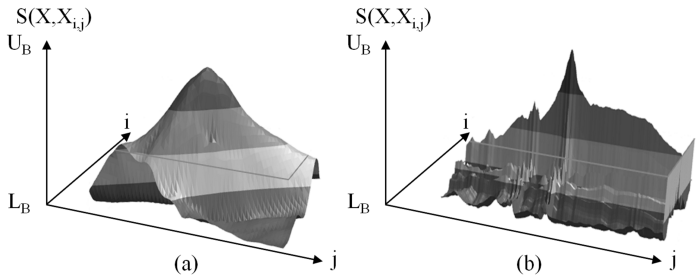

Figure 4. Peak with a low sharpness (i.e. smooth) (a) a high eccentricity (b)

B. Detection of the peak

One way to characterize robustness of a template to noise is to inspect the shape of the autocorrelation peak. Thus, we need to locate it first. A standard way is to threshold the autocorrelation map such as $W_{i,j} > \tau \ \forall i,j$. In [12], the threshold value τ is fixed by an expert user. We propose to determine τ from a performance characterization point of view, using the intra-class and between-class distributions.

We consider $X_1,\ldots,X_k,\ldots,X_{\complement}$ the set of templates and $Y_1,\ldots,Y_k,\ldots,Y_{\complement}$ a set of corresponding duplicate images. The intra-class corresponds to the $S(X_k,Y_k)\ \forall k$ comparisons, whereas the between-class is obtained from $S(X_k,Y_l)\ \forall k,l \neq k$. That is, the intra-class and between-class sets have cardinality of $\complement$ and $\complement(\complement-1)$ respectively (using no-commutative measures). We can obtain probability density functions from these sets such as $\int_{L_B}^{U_B} z_I(i) = \int_{L_B}^{U_B} z_B(i) = 1$. L_B, U_B are the lower and upper bounds of the considered similarity measure such as $L_B = S(X,\overline{X})$ and $U_B = S(X,X)$.

We have the separability gap Δ_S between the two distributions z_B, z_I as $\Delta_S = S_{max} - S_{min}$ with $S_{min} = \max(z_B)$ and $S_{max} = \min(z_I)$. If $\Delta_S > 0$ the matching problem is separable corresponding to a perfect detection score. With $\Delta_S < 0$, performance characterization enters in the Precision / Recall scheme. Considering the threshold $\tau \in [L_B, U_B]$, the true and false positive negative (tp, tn, fp, fn) are given with $tn = \int_{i=L_B}^{\tau} z_B(i)$, $tp = \int_{i=\tau}^{U_B} z_I(i)$ and $fp = 1 - tn$, $fn = 1 - tp$. Accuracy is then defined as $A = \frac{tp+tn}{tp+tn+fp+fn}$.

The threshold τ defines the precision of a pattern recognition system. Practical applications target to avoid any false positive fp and thus set τ high enough such as $\tau > S_{min}$. In our case, the challenge is to miss none peak in the localisation process and then we need to avoid any false negative fn. Thus, we setup the threshold low enough to guaranty $\tau < S_{max}$.

C. Peak characterization and robustness

Once the peak is located we can inspect the shape to characterize the robustness properties. In [12], sharpness is used to guaranty location accuracy while matching. We would want the shape of the peak as sharp as possible so that the position of the detected template will be accurate. The minimization of the sharpness will result in detection of smooth peaks as shown in Fig. 4 (a). The smooth peaks present robustness properties to noisy conditions. However, as we will discuss in our next section D this will tend to minimize the local derivatives then the goodness of templates for pruning.

We propose here to investigate eccentricity to characterize the peak robustness to noise. Eccentricity is correlated to sharpness but characterizes in addition the shape of the peak with elongation Fig. 4 (b). An elongated region appears as a specific sharp peak in the autocorrelation domain that can presents robustness properties to noisy conditions.

The eccentricity can be extracted by analysing the 2D shape of the peak using binary descriptors. Considering the set Q (i.e. $W_{i,j} > \tau \ \forall i,j$) we can compute the peak orientation θ_R and the eccentricity E_{CC}. The orientation θ_R describes the direction of the major axis, that runs through the peak center i_c, j_c. It can be found from the central moments μ_{pq} computation Eq. (2) with $\theta_R \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right[$. Similar to the orientation, the central moments can be used to determine the eccentricity E_{CC} Eq. (3). $E_{CC} \in [1, +\infty[$ that corresponds to a perfect circular disk $E_{CC} = 1$ and an elongated region $E_{CC} \gg 1$.

$$\mu_{pq} = \sum_{i,j \in Q} (i - i_c)^p (j - j_c)^q \quad \theta_R = \frac{1}{2} \arctan \frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \tag{2}$$

$$E_{CC} = \frac{\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}} \tag{3}$$

D. Autocorrelation features for pruning

As discussed in previous sections, the autocorrelation features can serve to characterize the robustness of templates to matching. We introduce here how they can be applied for matching optimization. Indeed, indexes can be obtained from an autocorrelation map and then used during the matching process to prune the search space. Here, the full autocorrelation map is processed to obtain the indexes and there is no need of a previous peak detection step (see Fig. 2).

Our approach is introduced in Fig 5. For elements $W_{i,j}$ $\forall i,j$, our approach looks for the maximum local derivatives $\Delta f(i)/\Delta i = (f(i+h) - f(i))/h$ in the horizontal direction, considering $f(i) = a$ a constant. $f(i)$ provides a $W_{i,j}$ value at the i,j position whereas $f(i+h)$, with $h \in [0, +\infty[$, is given with $W_{i+h,j} = U_B$. That is, $W_{i,j}$ is the closest peak element to $W_{i+h,j}$ having a value $f(i)$. Let us note that with $W_{i,j} = U_B$ we have $h = 0$ and then $\Delta f(i)/\Delta i \rightarrow +\infty$. A similar process is done for horizontal sampling to compute $f(j)$. The maximum local derivatives will ensure then a minimum sampling process to not miss the peak area.

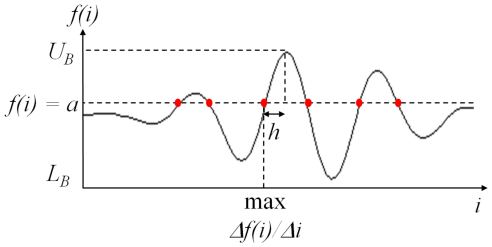

Figure 5. Introduction to features for pruning

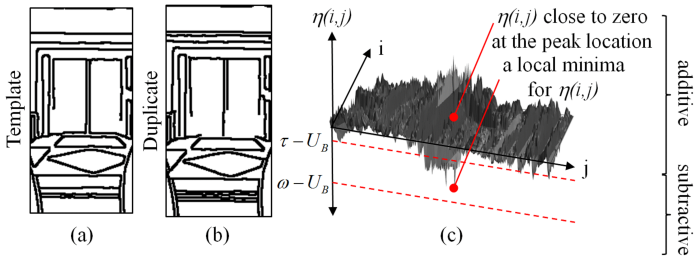

Figure 6. A template and a duplicate (a,b) the noise function $\eta(i,j)$ computed with the Yule measure $S(X,Y) \in [-1,1]$ [13]

The sampling parameters are recorded in arrays S^i, S^j. $S^i = (S^i_1, \ldots, S^i_k, \ldots, S^i_q)$ is a set of q quantified measures with $S^i_k \in [L_B, U_B]$ $\forall k$. At the beginning, we fix $S^i_k = \emptyset$ $\forall k$. We access with scan line the elements $W_{i,j}$ and obtain the corresponding k index with a lookup table function $k = LUT(W_{i,j})$. For elements $W_{i,j}$ $\forall i,j$, we set $S^i_k = \min(S^i_k, h)$ such as $h > 0$ and $W_{i+h,j} = U_B$. A similar process is done for horizontal sampling to compute the S^j.

The pruning must guaranty acceleration while preserving an exact matching result. To do this, the noise between the crosscorrelation and autocorrelation maps must be characterized. From Eq. (1), we define the crosscorrelation noise function $\eta(i,j) = S(X, Y_{i,j}) - S(X, X_{i,j})$ as shown in Fig. 6. In this function, $S(X, Y_{i,j})$ provides the similarity measure between the template X and the shifted duplicate image $Y_{i,j}$ and $S(X, X_{i,j})$ is obtained from the autocorrelation map. Thus, $S(X, Y_{i,j})$ and $S(X, X_{i,j})$ provide similarity measures at a same location i,j.

Two noise cases occur (1) the additive noise with $\eta(i,j) > 0$ and (2) the subtracting noise with $\eta(i,j) < 0$. The noise will result in offset values Δ_k when accessing S^i, S^j, with $\Delta_k > 0$ for additive noise and $\Delta_k < 0$ otherwise. To preserve the matching result, the Δ_k offsets should not result in oversampling i.e. $S^i_{k+\Delta_k} > S^i_k$ $\forall k$.

As a general trend, the S^i array appears as a decreasing function as the sampling parameters go down when converging to the peak area. We apply a min propagation to the S^i array $S^i_{k+1} = \min(S^i_k, S^i_{k+1})$ $\forall k$ that guaranties S^i to be a monotonically decreasing function. That is, we prevent oversampling with additive noise as we guaranty $S^i_{k+\Delta_k} < S^i_k$ with $\Delta_k > 0$. The process can be extended to S^j.

The maximum level of the subtracting noise can be known from a parameter $\omega \in [L_B, U_B]$ close to τ with $\omega < \tau$. τ is a global threshold that fixes the maximum level of noise at the peak location such as $S(X,Y) - S(X,X) > \tau - U_B$. $S(X,Y) - S(X,X)$ is a subset of the crosscorrelation noise and similar to τ, a threshold ω can be defined to fix $\eta(i,j) > \omega - U_B$ $\forall i,j$. Fig. 6 illustrates this aspect where (a) is a template and (b) is a duplicate. Their matching results in a low level of noise at the peak location $\eta(i,j) = -0.06$ whereas a local minima for the noise $\eta(i,j) = -0.22$ is observed on the rest of the map Fig. 6 (c).

The previous condition on $\eta(i,j), \omega$ can be rewritten in $|\eta(i,j)| < U_B - \omega$ $\forall i,j$. We can then apply a translation process to S^i with $T_k = LUT(U_B - \omega)$ and $S^i_k = S^i_{k-T_k}$ $\forall k$. That is, we prevent oversampling with subtracting noise as we guaranty $S^i_{k+\Delta_k} < S^i_k$ with $\Delta_k < 0$. A similar process can be applied to S^j. From the transformed S^i, S^j, the average wavelength for sampling λ is given in Eq. (4). The maximization of this feature characterizes the goodness of the template for pruning.

$$\lambda = \underset{\forall i,j}{\text{mean}}(d(LUT(W_{i,j})) \quad d(k) = \sqrt{(S^i_k)^2 + (S^j_k)^2} \qquad (4)$$

E. Fast matching

The complexity of the template matching depends of the image / template size parameters M, N, s, t and the used similarity measure. Since the template matching is unacceptably slow in most of the applications, a large effort has been made in the literature to design fast methods. A standard approach is to shift the image to a new representation space, where the matching can be accelerated for specific measures. For the binary similarity measures this includes the FFT, the RLE or the bitwise operators [14], [5]. In our approach we have considered the matching with bitwise operators as it is time efficient for local matching, supports the computation with all the measures and can be driven while pruning.

The general algorithm for matching while pruning, using S^i, S^j indexes, is detailed thereafter. R is the region of interest where the matching is applied. We have $R \in W$ and having a size $n' = s' \times t'$ with $s' < 2s+1$, $t' < 2t+1$. B is a boolean matrix of size $n' = s' \times t'$ and $B_{i,j}$ its set of elements. At the initialization, we have $B_{i,j} = 0$ $\forall i,j$. Acceleration resulting of this algorithm is obtained with $\varpi = n'/(n' - \sum_{\forall i,j} B_{i,j})$. The matching algorithm outline is then:

- To encode R for matching with bitwise operators.
- At every pixel location $(i,j) \in B$, if $B_{i,j} = 0$ compute $S(X, Y_{i,j})$ with operators [14], [5].
- Then, get the corresponding k index with the LUT function and do $B_{i+y,j+x} = 1$ $\forall y \in [0, S^i_k[$ and $\forall x \in [0, S^j_k[$, set $B_{i,j} = 0$.

F. Selection process

At last, from the autocorrelation features we must specify a rule to drive the selection process. We assume different hypothesis to drive the selection:

• The λ, E_{CC} feature sets are close to normal distributions and their maximization is correlated.
• The extremum detection with average wavelength λ guaranties a high value range for the acceleration factor ϖ.

• The extremum detection with peak eccentricity E_{CC} guaranties a high value range for $L^{max} = S(X, Y)$.

Considering these hypothesis, the selection becomes an outlier detection problem. We propose a two-steps strategy to ensure optimization then robustness with selection. We define σ_i, σ_j the standard deviations of the λ (i), E_{CC} (j) distributions with their centroids μ_i, μ_j. With a normal distribution, a standard rule for outlier detection is to consider the conditions (a) $\lambda > \mu_i + 3\sigma_i$ (b) $E_{CC} > \mu_j + 3\sigma_j$. We detect outliers by respecting $a \bullet b$, with $\bullet$ the logical AND operator. If several templates are detected as outliers, we finalize the process with maximization of the λ feature.

IV. Experiments and results

In this section we present experiments and results about our approach. We will introduce in section A the dataset used in our experiments. Section B will characterize our template selection process, whereas section C will present performance in terms of matching results and time processing.

A. Dataset

At best of our knowledge, the exact duplicate detection problem of Manga / Comics images has neither been addressed in the literature. Our first task was to constitute a dataset we will refer as $MangaOPU$. We have constituted an image dataset composed of $3844 \times 2 = 7688$ legal and illegal pages of a Manga Magazine - *Manga Shukan Shonen Jump*[1]. The illegal images have been collected from web portals. They are given at low resolution (128 dpi with a mean page size of 1300×900 pixels), compressed using the jpg standard with a low quality factor and produced with homemade digitalization processes. Legal images have been produced at high resolution and quality in terms of compression and digitalization process, and then downsampled for comparison. The binary images have been obtained using a system described in [5] with gray-level conversion and canny-edge detection.

B. Performance characterization for selection

To characterize our selection process, a first step is to establish a correlation between the autocorrelation features and the matching results. To do this, we need to constitute an initial set of "good" template. To tackle this problem, we have first driven a selection in a reference context as did in [9], [11], [12]. We have extracted randomly 30 templates 256×128 per page and applied them for matching. For our experiments, we have driven the matching with the weighted Yule measure [13]. We have then selected the "best" template per page corresponding to the strongest local maxima $L^{max} = S(X, Y)$. We have characterized then the existing correlation between our autocorrelation features and the L^{max} values. The complete set of experiments is presented in Figures (7) (a-f).

In Figures (7) (a-c), peaks are characterized according their eccentricity E_{CC} and orientation θ_R.

Figure (7-a) provides the E_{CC} distribution using a dB scale. Peaks appear mainly ($\simeq 95\%$) as a near-blob structure with low level $E_{CC} < 2$ dB. Normality test reveals that

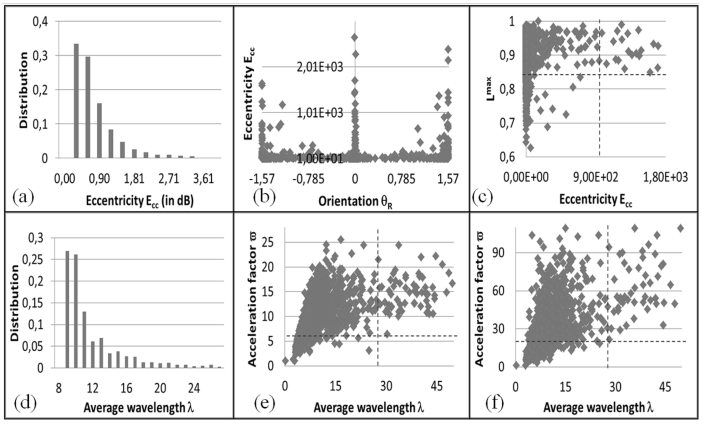

Figure 7. Characterization of the template selection (a-c) peak eccentricity and orientation (d) average wavelength distribution (e) average wavelength / acceleration - the tp case (f) average wavelength / acceleration - the tn case

the E_{CC} distribution is close to a normal distribution as highlighted in Figure (7-a). Within the range $E_{CC} \in [0, 2[$ dB, the orientation parameter θ_R is little accurate as a blob does not run through a main direction axis. Therefore, we have $\theta_R \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ for $E_{CC} < 2$ dB. With highest $E_{CC} > 2$ dB, some main orientations appear $|\theta_R| \simeq \{0, \frac{\pi}{2}\}$ Figure (7-b).

Figure (7-c) evaluates the correlation between L^{max} and E_{CC}. More a peak converges to a blob structure, more it becomes sensitive to lowest L^{max}. From this distribution, the selection threshold $E_{CC} > \mu_j + 3\sigma_j$ is reported in dot line in Figure (7-c). With $E_{CC} > 936.7$ (i.e. 2.97 dB) we have $L^{max} \in [0.84, 1]$ with a mean value $L^{max} = 0.92$.

Figures (7) (d-f) discuss correlation between the average wavelength λ and the acceleration factor ϖ.

Figure (7) (d) highlights a normal distribution for λ checked with a normality test. Figure (7) (e) presents the correlation between λ and ϖ. More λ increases, better the acceleration factor ϖ is. The selection threshold $\lambda > \mu_i + 3\sigma_i$ is reported in dot line. For $\lambda > 27.7$ we obtain $\varpi \in [6, 34]$ with a mean value $\varpi = 15.08$. We have extended this analysis to the true negative tn case Figure (7) (f). We have computed the between-class distribution $S(X_k, Y_l)$ $\forall k, l \neq k$ with our template set and look for pruning abilities. For $\lambda > 27.7$ we obtain $\varpi \in [31, 265]$ with a mean value $\varpi = 90.72$. This is the major pruning result as the recognition process drives $\mathsf{C}-1$ tn comparisons and 1 tp comparison from a dataset of size C.

C. Performance characterization for matching

Table I provides information about processing time[2]. The matching time depends mainly on the $s' \times t'$ parameters for R. Normality tests reveal that the image registration parameters (i.e. i, j location differences between X and Y) are close to normal distributions. Within the $MangaOPU$ dataset, 99% of the registration cases appear in $-3\sigma, +3\sigma$ whereas a full coverage is obtained with $-5\sigma, +5\sigma$. In that case, we obtain with experiments $s' \times t' = 64 \times 128$.

For matching with bitwise operators [14], [5] we need to encode (i.e. get the fingerprint) the region of interest R first.

[1] N° 26, 27, 28, 35, 41, 42, 44, 45, 46 and 48

[2] Java/C++ framework on a Windows 7 system with an 2.1 GHz Intel CPU

method	fingerprint	matching	
Our approach	$0.37\ ms$	FS	$16.4\ ms$
		FS with pruning (tp)	$1.08\ ms$
		FS with pruning (tn)	$0.18\ ms$
GIST	840 ms		0.003 ms

Table I. COMPARISON OF PROCESSING TIME

This requires then some hundreds μs. Then, comparing a 256×128 template X to a subimage Y at a location i, j can be achieved in 2 μs. Considering $s' \times t' = 64 \times 128$, the FS application of our algorithm requires $16.4\ ms$ for the complete matching. Wile pruning, the processing time goes down to 1 ms for a tp detection and 180 μs for fp as averages. Our full processing time (encoding then matching) is near to a half ms in the main (i.e. fn) case. Using our autocorrelation indexes S^i, S^j, we have fixed a setting $\omega = 0.12$, $\tau = 0.22$ that preserves exact matching results while pruning.

With the template set, we have obtained separable distributions $\Delta_S > 0$, $A = 1$ on the dataset (i.e. separability). Fig. 8 gives some of the worst detected cases. We have completed characterization with a validity indice. A validity indice helps to complete characterization on a separable problem. We have used the Davies - Bouldin Index DBI as a validity indice of the z_B, z_I distributions. Considering a two-class problem i, j, the DBI could be given with the standard deviations σ_i, σ_j as dispersion measure and $d_{i,j} = \mu_i - \mu_j$ the distance between the two centroids. Thus, the indice is given as $DBI = (\sigma_i + \sigma_j)/d_{i,j}$. The indice operates in the range $DBI \in [0, +\infty[$, with $DBI = 0$ the optimal value and $[0, \frac{1}{3}]$ the usual range for separable problems[3]. Our obtained DBI is $11,714 \times 10^{-3}$ close to the optimal value $DB = 0$.

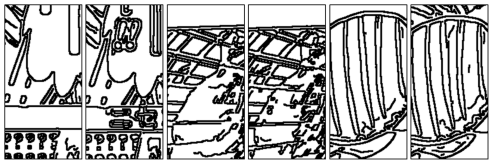

Figure 8. Some worst detected cases (left) images (right) templates

For comparison, we have used the GIST descriptor considered as a reference method for exact duplicate detection in the Computer Vision field [1], [3]. We have set GIST for optimum performance by processing the 300×300 central part of gray-level images in a not-scale and not-rotation invariant way for fair comparison. We have obtained separability but with a higher $DBI = 288,627 \times 10^{-3}$. This indice is close to the separability upper bound $\frac{1}{3} = 333,333 \times 10^{-3}$. That is GIST performances are worst and our approach is supposed to preserve separability when faced to largest recognition problems whereas GIST should not.

Another major difference is processing time Table I. GIST requires a large processing time for fingerprint extraction, near to $1s$ to process a 300×300 image. At the matching level, comparing two 1296 GIST descriptors can be done in 3 μs. Thus, GIST scales better when $\mathtt{C}$ is large. Our approach can process one to two thousands template to image comparisons while GIST extracts a fingerprint. GIST looks more suitable for retrieval, whereas our approach presents better performance for recognition or when real-time constraints are needed.

[3]Considering a coverage of 6σ for a Gaussian distribution.

V. CONCLUSION AND PERSPECTIVES

We have presented an exact duplicate image detection method based on binary template selection and matching. Through the selection, the matching is made robust and time efficient. The approach has been tested on a proposed dataset against the GIST descriptor. We obtain better performances in terms of robustness and response time. We plan to extend the experimental aspects. The Manga / Comics duplicate detection is a recent topic of interest [4], [5], [6] and none public dataset exists. Recently, a Manga dataset[4] has been made available for research. This requires an image degradation framework to get the synthetic near-duplicates. Complete performance characterization of our method must be driven against the literature methods [1], [3], [2]. At last, further acceleration of our approach could be obtained with FS-equivalent matching and template search space reduction technics.

REFERENCES

[1] X. Wang, L. Zhang, and C. Liu, "Duplicate discovery on 2 billion internet images," in *Workshops on Computer Vision and Pattern Recognition (CVPRW)*, 2013, pp. 2160–7516.

[2] L. Liu, Y. Lu, and C. Suen, "Near-duplicate document image matching: A graphical perspective," *Pattern Recognition (PR)*, vol. 47, no. 4, pp. 1653–1663, 2014.

[3] Y. Lei, L. Zheng, and J. Huang, "Geometric invariant features in the radon transform domain for near-duplicate image detection," *Pattern Recognition (PR)*, vol. 47, no. 11, pp. 3630–3640, 2014.

[4] S. Weihan and K. Kise, "Detection of exact and similar partial copies for copyright protection of manga," *International Journal on Document Analysis and Recognition (IJDAR)*, pp. 1–19, 2013.

[5] M. Delalandre, M. Iwata, and K. Kise, "Fast and optimal binary template matching, application to manga copyright protection," in *Workshop on Document Analysis Systems (DAS)*, 2014, pp. 298–303.

[6] T. Le, M. Luqman, J. Burie, and J. Ogier, "Retrieval of comic book images using context relevance information," in *Workshop on coMics ANalysis, Processing and Understanding (MAPU)*, 2016, pp. 12:1–12:6.

[7] A. Adnan, "Optimum template selection for image registration using icmm," in *British Machine Vision Association Conference (BMVC)*, 1998, pp. 1–11.

[8] A. Bazen and al, "A correlation-based fingerprint verification system," in *Workshop on Circuits, Systems and Signal Processing (ProRISC)*, 2000, pp. 205–213.

[9] U. Uludag, A. Ross, and A. Jain, "Biometric template selection and update: a case study in fingerprints," *Pattern Recognition (PR)*, vol. 37, pp. 1533–1542, 2004.

[10] T. Han, V. Ramesh, Y. Zhut, and T. Huang, "On optimizing template matching via performance characterization," in *International Conference on Computer Vision (ICCV)*, vol. 1, 2005, pp. 182–189.

[11] M. Debella-Gilo and A. Kaab, "Locally adaptive template sizes for matching repeat images of mass movements," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2011, pp. 4281–4284.

[12] H. Choi, R. Gupta, and S. Suh, "Quality measurement of template models and automatic template model selection," in *International Conference on Control, Automation and Systems (ICCAS)*, 2012, pp. 1044–1048.

[13] S. Choi, S. Cha, and C. Tappert, "A survey of binary similarity and distance measures," *Journal of Systemics, Cybernetics and Informatics (SCI)*, vol. 8, no. 1, pp. 43–48, 2010.

[4]http://www.manga109.org/

[14] S. Mukherji, "Fast algorithms for binary cross-correlation," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 1, 2005, pp. 340–343.