# A Local Adaptation of the Histogram Radon Transform Descriptor: An Application to a Shoe Print Dataset

Makoto Hasegawa and Salvatore Tabbone

Université de Lorraine-LORIA, UMR 7503,
54506 Vandoeuvre-lès-Nancy, France
{makoto.hasegawa,tabbone}@loria.fr

**Abstract.** In this paper we propose a shape recognition approach applied to a dataset composed of 512 shoeprints where shapes are strongly occluded. We provide a local adaptation of the HRT (Histogram Radon Transform) descriptor. A shoeprint is decomposed into its connect components and describes locally by the local HRT. Then, following this description, we find the best local matching between the connected components and the similarity between two images is defined as mean of local similarity measures.

**Keywords:** Shape matching, local descriptor, histogram of Radon transform.

## 1   Introduction

In pattern recognition, image descriptions can be broadly categorized into statistical methods or structural methods. In the statistical method, many methods for a whole image have been proposed in the literature. The generic Fourier descriptor (GFD) proposed by Zhang and Lu [1] is a typical Fourier descriptor. The Fourier-Mellin transform (FMT) proposed Q. Chen et al. [2] is useful for many applications. Tabbone et al. propose methods called the histogram of Radon transform (HRT) [3] using the Radon transform. Recently, a local descriptor with vector of feature points in an image called SIFT is proposed by Lowe [4]. Structural methods[5] offer a good description thanks to the graph representation but they are not robust to noise.

In this paper, we consider a shoeprint dataset where shoeprints can be decomposed into connected components.The shoeprint can be decomposed into some connected components using the connectivity of 8 pixels around each pixel and each connected component is encoded using HRT for a local description. HRT has useful properties for shape rotation, shape scaling, and shape translation; then it is robust to geometric transformations of components. For a query shoe print image, we find the best local matching between the connected components of the query and the dataset and the similarity between two images is defined as mean of local similarity measures.

We have carried out our experiments using a dataset composed of 512 shoeprints and we will show that our approach is competitive compared to well-known statistical descriptors (global HRT and FMT) and structural (graph edit distance) one.

## 2    HRT Descriptor

We recall the Radon transform definition in this section. Let a coordinate $(x, y)$ in the two-dimensional $x - y$ plane described as $\mathbf{x}$, and an original image represented as $f(\mathbf{x})$. The Radon transform of $f(\mathbf{x})$ is defined as:

$$\mathcal{R}_f(\theta, \rho) = \int f(\mathbf{x})\delta(\mathbf{x} \cdot \xi - \rho)\mathrm{d}\mathbf{x}, \tag{1}$$
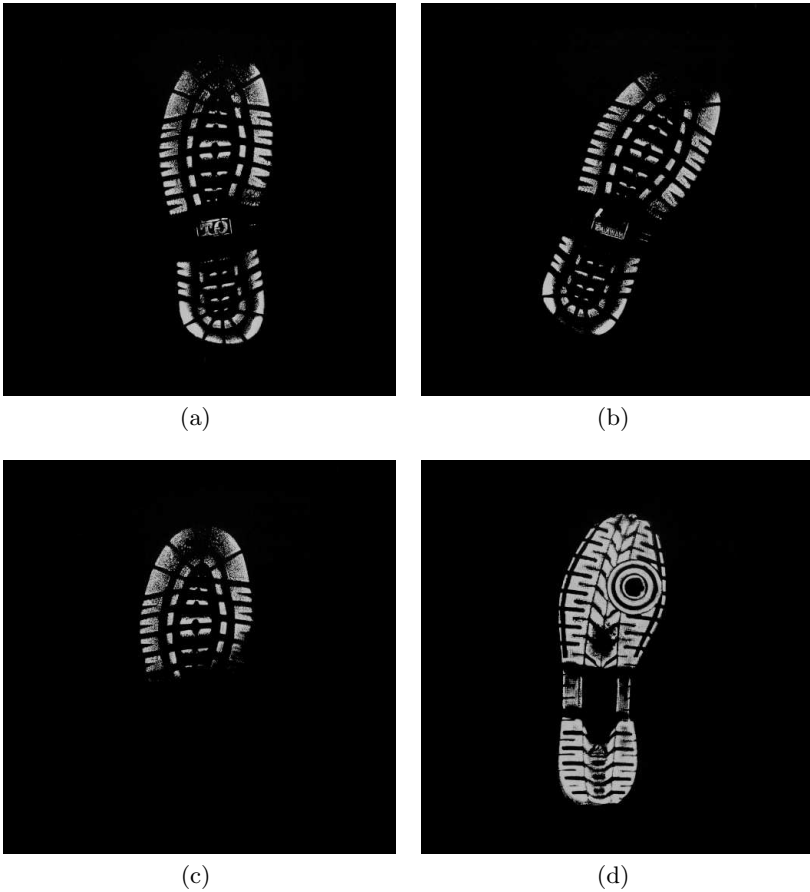


**Fig. 1.** (a) Full-print of "Shoe A"; (b) rotation; (c) occlusion. (d) Full-print of "Shoe B".
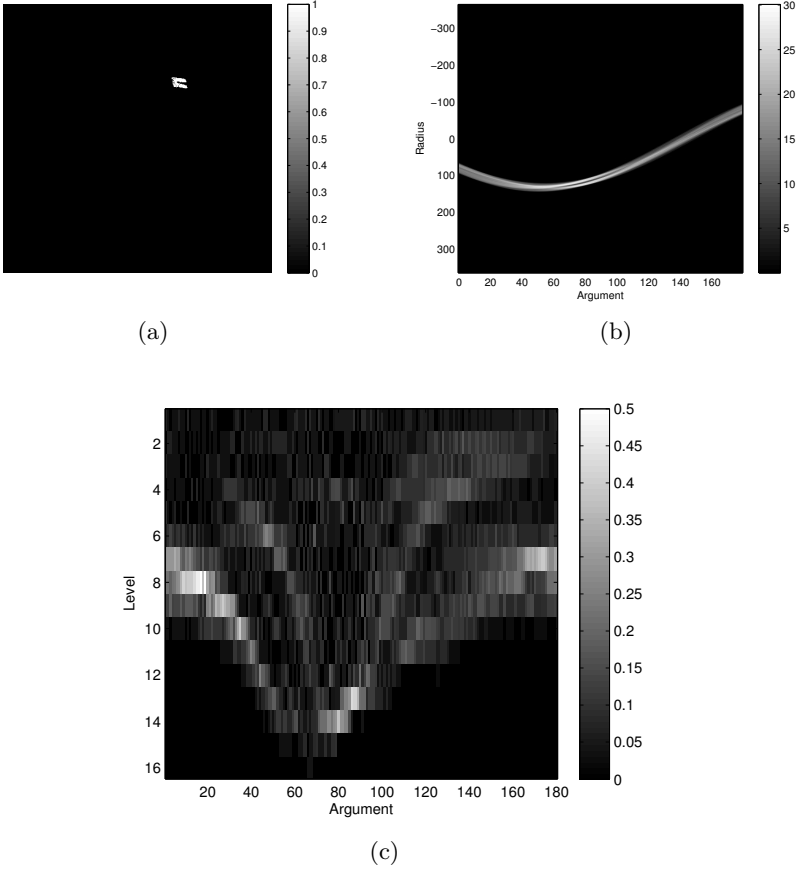
(a)



(b)



(c)

**Fig. 2.** (a) A connected component of "Shoe print A"; (b) and (c) respectively its Radon image and HRT descriptor

where $\xi = (\cos\theta, \sin\theta)$, and $\delta(\cdot)$ is a delta function. In other words, the Radon transform is the integral of $f(\mathbf{x})$ over lines

$$L_{\theta\rho} = \{\mathbf{x} \in R^2 | \mathbf{x} \cdot \xi = \rho\}, \tag{2}$$

where $\rho$ is the distance between the origin and $L_{\theta\rho}$, the unit vector $\xi$ and the angle $\theta$ describe the orientation of the line $L_{\theta\rho}$. The line integral is computed by a delta function $\delta(\cdot)$. An example of the Radon image using Fig. 2(a) is shown in Fig. 2(b); Fig. 2(a) is a connected component of "Shoe A" shown in Fig. 1(a).

The HRT descriptor is defined as a matrix of frequencies computed on the Radon image aggregated by the angle parameter of the Radon transform. The HRT descriptor $D(\theta, y)$ of $\dot{R}_f(\theta, \rho)$ for each orientation $\theta$ is:

$$D(\theta, y) = H(\dot{R}_f(\theta, \cdot))(y), \tag{3}$$

where $H(f)(y)$ is a histogram of $f$ as:

$$H(f)(y) = \frac{\#\{x \in X | y = f(x)\}}{|X|}. \tag{4}$$

$\#$ in this equation is the cardinality of a set, and $|X|$ is the cardinality of the universal set. The universal set $X$ is composed of elements in the case of $f(x) > 0$ as: $X = \{x | f(x) > 0\}$. The number of bins in the histogram is defined by experiments and we set it to 16 since it provide us the best performance.

$\dot{R}_f(\theta, \rho)$ in Eq.(3) is a normalized Radon image; here a normalization is defined as:

$$\dot{R}_f(\theta, \rho) = \frac{R_f(\theta, \rho)}{N}, \tag{5}$$

where

$$N = \max R_f(\theta, \rho). \tag{6}$$

Our HRT descriptor has useful properties for shape rotation, shape scaling, and shape translation. In the shape rotation, our HRT descriptor is translated relative to the angle coordinate; in the shape scaling and translation, our HRT descriptor is invariant.

$D_q(\theta, y)$ denotes a query descriptor, and $D_t(\theta, y)$ is for a template; we define a matching error as:

$$E(D_q, D_t) = \min_{\alpha \in [0, \pi)} \frac{\sum_{y, \theta} (D_q - \bar{D}_t)(D_q - \bar{D}_t)}{\sqrt{(\sum_{y, \theta}(D_q - \bar{D}_t)^2)(\sum_{y, \theta}(D_q - \bar{D}_t)^2)}}, \tag{7}$$

where $D_q = D_q(\theta, y)$, $\bar{D}_q$ is the means of $D_q$, and $\sum_{y, \theta} = \sum_y \sum_\theta$.

## 3   Extension to Shape Matching Using Local Descriptor

We decompose shapes on images into its connected components using the connectivity of 8 pixels around each pixel. Fig. 3 shows examples of the decomposition for shoeprint images. Each color of the connected component shows its component label.

Subsequently, each connected component is encoded using HRT, and we obtain two sets of descriptors of components, one is for a query, and the other is for a template.

We perform our local matching between the query descriptors and the template and find the closest one to each query component from the template components so that each best matching pair between the query and the template is defined. Matching error for each component pair is computed by Eq. (7); a mean value of the best matching pairs is the matching error between two images. Fig. 4 shows examples of our local shape matching between a query and a shoeprint image. We can remark that in some cases the matching between two connected components is not coherent and spatial constraints should be added to improve it.
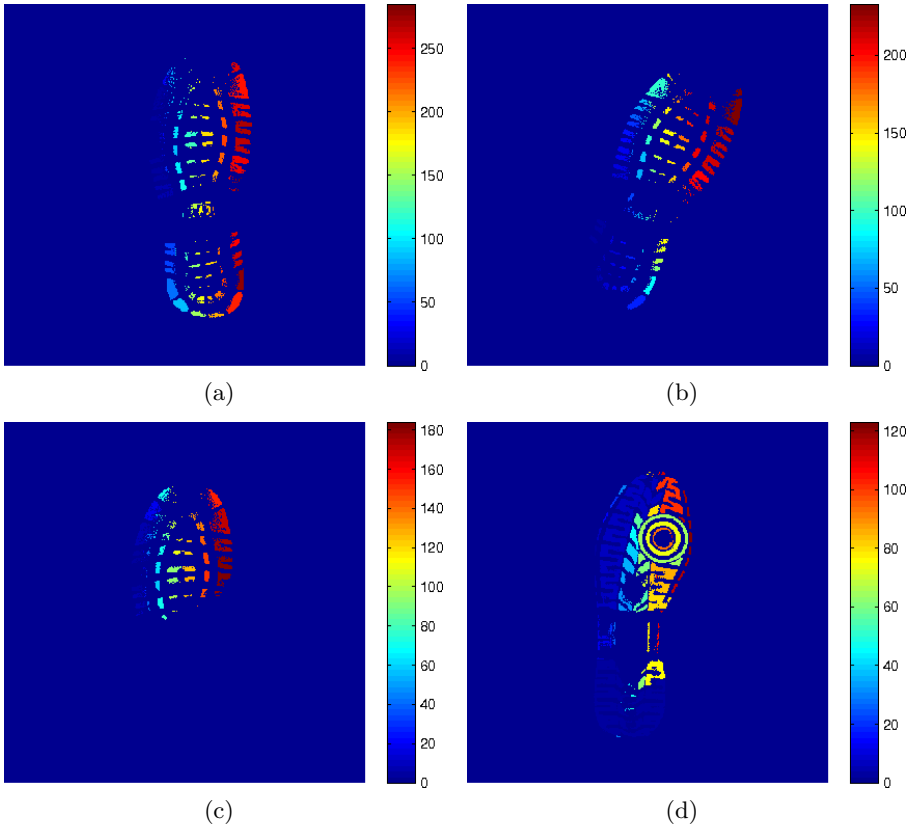
**Fig. 3.** Connected component decomposition

## 4 Experimental Results

The shoeprint dataset composed of 512 shoeprints; each shoeprint is a binary image with $512 \times 512$ pixels. Different example of shoes are shown in Fig. 5; each shoe has 8 images (4 images are full-prints; 4 images are strongly occluded: only toe, heel, light side, and left side). We will provide our dataset for a public common use.

### 4.1 Recognition Performance

Examples of matching score of our local descriptor HRT are shown in Table 1. We can see that shoes of the same class A are nearer than shoe B. The performance on the whole dataset is performed using the precision-recall curves defined as:

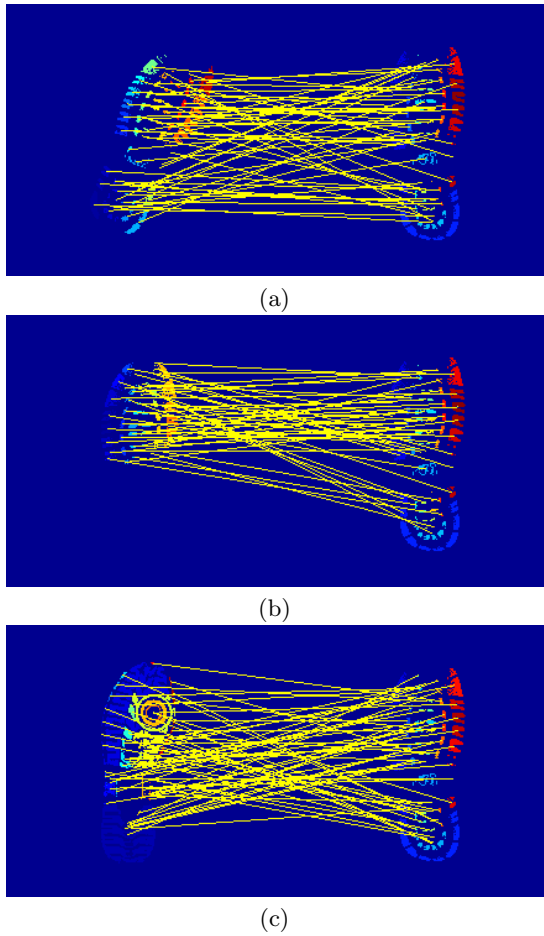$$\text{Precision} = \frac{tp}{tp + fp}, \quad \text{Recall} = \frac{tp}{tp + fn}, \tag{8}$$

(a)

(b)

(c)

**Fig. 4.** Example of local shape matching using HRT local descriptors

**Table 1.** Matching errors using HRT local descriptor

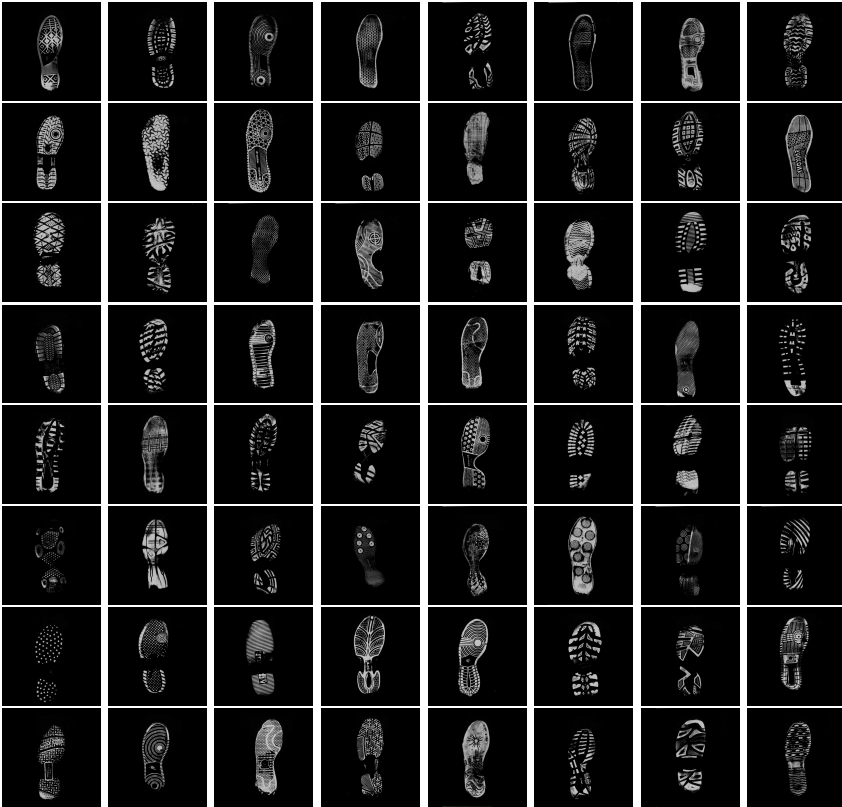|  |  |  | Template | | | |
|---|---|---|---|---|---|---|
|  |  |  | Shoe A | | | Shoe B |
|  |  |  | Fig.1(a) | Fig.1(b) | Fig.1(c) | Fig.1(d) |
| Query | Shoe A | Fig.1(a) | 0.00 | 142.55 | 157.83 | 179.20 |
|  |  | Fig.1(b) | 140.71 | 0.00 | 161.81 | 170.98 |
|  |  | Fig.1(c) | 148.34 | 152.97 | 0.00 | 181.17 |
|  | Shoe B | Fig.1(d) | 182.62 | 179.15 | 196.05 | 0.00 |

**Fig. 5.** Shoeprint dataset

where $tp$ is the number of items correctly labeled as belonging to the positive class, $fp$ is the number of items incorrectly labeled as belonging to the class, and $fn$ is the number of items which are not labeled as belonging to the positive class but should have been.

We compare the precision-recall curves of our methods to the conventional HRT applied to a whole image, the Fourier-Mellin transform (FMT) and graph edit distance; the results are shown in Fig. 6. We can remark that global approaches are not suited for this kind of dataset and the performance provided by our local descriptor are globally good even if images are strongly occluded. For instance, as we can see in Table 1 the occluded shoeprint on Fig. 1(c) is nearer to Fig. 1(a) and 1(b) than Fig. 1(d).

## 4.2   Processing Time

Processing time for one matching using each method is shown in Table 2. Each method is performed by Intel Core i5-460 2.53 GHz CPU. HRT for local descriptor, Conventional HRT, and FMT are implemented using MATLAB and the
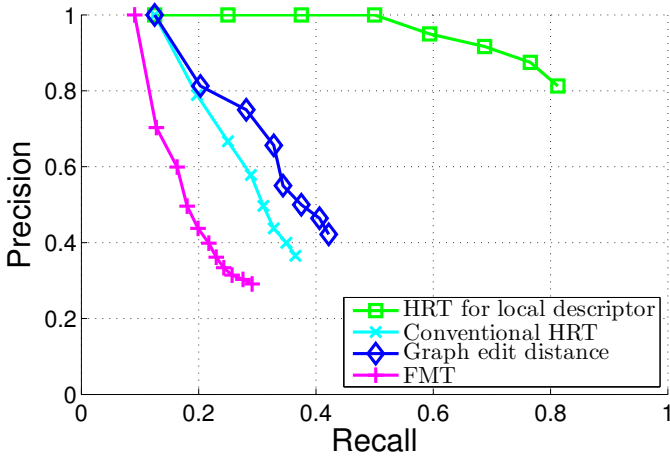
(a)

**Fig. 6.** Precision-recall curves of each method

**Table 2.** Processing time for one matching

| Method | Processing time (Sec) |
|---|---|
| HRT for local descriptor | 0.825 |
| Conventional HRT | 0.014 |
| Graph edit distance | 6.998 |
| FMT | 0.036 |

graph edit distance code provided in [5] is in C++. The processing time for our HRT local descriptor is high compared to the conventional HRT and FMT since we apply it locally.

## 5    Conclusion

We discuss a shape matching method using local descriptor. We apply HRT to connected components for local descriptions and find the best local matching between connect components. The obtained results are promising and could be improved by achieving a better decomposition of the image. Indeed, the performance of our local descriptor is strongly depend on the quality of the segmentation. In our case we use a simple decomposition based on the connected component. However, in some cases this decomposition is not appropriate. For instance, Fig. 7 show an original shape "Shoe B" and an connected component extracted on this image. This connected component has not been sufficiently decomposed, and in some cases a shoeprint is decomposed into one component only. This decomposition error leads to recognition errors and decreases the
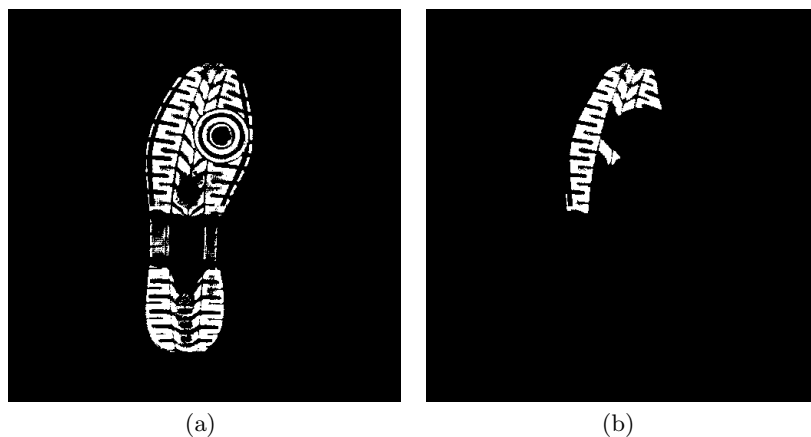
Fig. 7. (a) Original shape "Shoe B"; (b) A large connected component

recognition performance. In this perspective, future investigations will be devoted to an appropriate shape decomposition method.

## References

1. Zhang, D., Lu, G.: Shape-based image retrieval using generic Fourier descriptor. Sig. Proc.: Image Comm. 17(10), 825–848 (2002)
2. Sheng Chen, Q., Defrise, M., Deconinck, F.: Symmetric phase-only matched filtering of Fourier-Mellin transforms for image registration and recognition. IEEE Trans. PAMI 16(12), 1156–1168 (1994)
3. Tabbone, S., Ramos Terrades, O., Barrat, S.: Histogram of Radon transform. A useful descriptor for shape retrieval. In: International Conference on Pattern Recognition, Tampa, pp. 1–4 (2008)
4. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
5. Riesen, K., Bunke, H.: Approximate graph edit distance computation by means of bipartite graph matching. Image Vision Comput. 27(7), 950–959 (2009)