Feature description based on Mean Local Mapped Pattern

Carolina Toledo Ferraz∗, Osmando Pereira Junior†, Adilson Gonzaga‡
Department of Electrical Engineering - EESC/USP
University of São Paulo
Av. Trabalhador São Carlense, 400
13560-590, São Carlos, SP, Brasil
∗caroltoledoferraz@gmail.com, †osmandoj@gmail.com, ‡agonzaga@sc.usp.br

Abstract—Local feature description has gained a lot of interest in many applications, such as texture recognition, image retrieval and face recognition. This paper presents a novel method for local feature description based on gray-level difference mapping, called Mean Local Mapped Pattern (M-LMP). The proposed descriptor is robust to image scaling, rotation, illumination and partial viewpoint changes. Furthermore, this descriptor more effectively captures the nuances of the image pixels. In our experiments, the descriptor is compared to the Center-Symmetric Local Mapped Pattern (CS-LMP) and the Center-Symmetric Local Binary Pattern (CS-LBP). The results show that our descriptor performs better compared to these two methods.

I. INTRODUCTION

In image processing, the local feature description plays an important role in applications such as texture recognition, content-based image retrieval and face recognition. The objective is to build a feature vector providing a representation that allows efficient matching of local structures between images.

One of the most widely known key feature descriptors published in the literature is the scale-invariant feature transform (SIFT) [1]. It was introduced by Lowe in 1999 [2]. It is characterized by a 3D histogram of gradient locations and orientations and stores the bins in a vector of 128 positions. Many other descriptors like SIFT have been proposed, such as Principal Components Analysis SIFT (PCA-SIFT) [3], Gradient Location and Orientation Histogram (GLOH) [4], Speeded Up Robust Features (SURF) [5], Colored SIFT (CSIFT) [6], Center-Symmetric Local Binary Pattern (CS-LBP) [7] and Kernel Projection Based SIFT (KPB-SIFT) [8]. The GLOH descriptor is very similar to SIFT, however, it uses a Log-Polar location grid instead of a Cartesian one. SURF approximates SIFT using the Haar wavelet response and using integral images to compute the histograms bins. It uses a descriptor vector of 64 positions, providing a better processing speed. Jin, Liu, Lu and Tong [9] proposed the Improved Local Binary Pattern (ILBP) that is an improvement of the LBP feature and is used for face detection [9]. It compares all of the pixels (including the central pixel) with the mean intensity. On the other hand, the Mean Local Binary Pattern (M-LBP) descriptor [10] does not consider the central pixel in this calculation. Center-Symmetric Local Binary Pattern (CS-LBP) [7] combines the strengths of the well-known SIFT descriptor and the Local Binary Pattern (LBP) texture operator. In [7], it is shown that the construction of the CS-LBP descriptor is simpler than SIFT, but it generates a feature vector of 256 positions, which is twice the SIFT vector size.

In the Local Mapped Pattern (LMP) approach [11], the authors consider the sum of the differences of each gray-level of a given neighborhood to the central pixel as a local pattern that can be mapped to a histogram bin using a mapping function.

The Center-Symmetric Local Mapped Pattern (CS-LMP) [11] combines the desirable properties of the CS-LBP and the Local Mapped Pattern (LMP). This approach is based on the sum of the differences of each gray-level of a given neighborhood to the center-symmetric pairs of pixels as a local pattern that can be mapped to a histogram bin using a mapping function.

In this paper, we propose a new interest region descriptors, denoted as the Mean Local Mapped Pattern (M-LMP). The M-LMP descriptor is based on the CS-LBP, ILBP and M-LBP using the LMP methodology. This new descriptor captures the small transitions of pixels in the image more accurately, resulting in a greater number of correct matches than the CS-LMP and CS-LBP.

The rest of this paper is organized as follows. In Section II, we briefly describe the CS-LMP, CS-LBP, ILBP, M-LBP and LMP. Section III gives details on the proposed approach. The experimental evaluation and the results are presented in Section IV. Finally, we conclude the paper in Section V.

II. LOCAL DESCRIPTORS

In this section, five previously published local descriptors are shown as the theoretic material for our approach: the CS-LBP, CS-LMP, ILBP, M-LBP and LMP.

A. Center-Symmetric Local Binary Pattern (CS-LBP)

The CS-LBP is a modified version of the LBP texture feature and SIFT descriptor. In [7], the authors state that the LBP produces a rather long histogram and therefore is difficult to use in the context of an image descriptor. To solve this problem, they modified the scheme of how to compare the pixels in the neighborhood. Instead of comparing each pixel with the center one, the method compares center-symmetric pairs of pixels.

For the construction of the descriptor, the image regions are divided into cells with a location grid, and for each cell, a
CS-LBP histogram is built. The CS-LBP feature extraction for each pixel of the region is accomplished according to Equation 1:

\[
CS - LBP_{R,N,T}(x,y) = \sum_{i=0}^{(N/2)-1} s(n_i - n_{i+(N/2)})2^i,
\]

(1)

where \(n_i\) and \(n_{i+(N/2)}\) correspond to the gray values of center-symmetric pairs of pixels of N equally spaced pixels on a circle of radius R and

\[
s(x) = \begin{cases} 
1, & \text{if } x > T, \\
0, & \text{otherwise}. 
\end{cases}
\]

\(T\) is set to 0.01 [12]. For a 4 × 4 Cartesian grid, the final feature vector contains 256 positions.

**B. Local Mapped Pattern (LMP)**

The LMP methodology assumes that each gray-level distribution within an image neighborhood is a local pattern. This pattern can be represented by the gray-level differences around the central pixel [11]. Figure 1 shows an example of a 3 × 3 neighborhood and a 3D graphic of the respective local pattern.

![Fig. 1. Gray-level differences as a local pattern](image)

Each pattern defined by a \(W \times W\) neighborhood will be mapped to a histogram bin \(h_b\) using Equation 2, where \(f_{g(i,j)}\) is the mapping function, \(P(k,l)\) is a weighting matrix of predefined values for each pixel position within the neighborhood, and \(B\) is the number of the histogram bins. This equation represents the weighted sum of each gray-level difference between the neighboring pixels and the central one, mapped onto the [0,1] interval by a mapping function and rounded to \(B\) possible bins.

\[
h_b = \text{round} \left( \frac{\sum_{k=1}^{W} \sum_{l=1}^{W} f_{g(i,j)} P(k,l)}{\sum_{k=1}^{W} \sum_{l=1}^{W} P(k,l)} \right) (B - 1),
\]

(2)

For texture analysis, the LMP approach uses a common sigmoid curve as in Equation 3.

\[
f_{g(i,j)} = \frac{1}{1 + e^{-[A(k,l) - g(i,j)]}},
\]

(3)

where \(\beta\) is the curve slope and \([A(k,l) - g(i,j)]\) are the gray-level differences within the neighborhood centered in \(g(i,j)\). The proposed weighting matrix is:

\[
P(k,l) = \begin{bmatrix} 
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1 
\end{bmatrix}
\]

**C. Center-Symmetric Local Mapped Pattern (CS-LMP)**

The CS-LMP is based on the CS-LBP, but using a mapping function in the LMP method [11]. Equation 4 shows the feature extraction for each pixel of the interest region, defined in a neighborhood \(W \times W\):

\[
CS - LMP_{R,N,f}(x,y) = \text{round} \left( \frac{\sum_{i=0}^{N/2-1} f(i)P(i)}{\sum_{i=0}^{N/2-1} P(i)} \right) b
\]

(4)

where \(b\) is the number of histogram bins minus one, \(P(i)\) is the \(i\)th element of the weight matrix \(P\) defined in the Equation 5, and \(f(i)\) is the sigmoid mapping function defined in Equation 6.

\[
P = [p_0,p_1,...,p_{N/2-1}]^T
\]

(5)

\[
f(i) = \frac{1}{1 + e^{-[A_{i} - g(i,j)]/\beta}}
\]

(6)

where \(n_i\) and \(n_{i+(N/2)}\) correspond to the gray level values of the center-symmetric pairs of \(N\) equally spaced pixels on a circle of radius \(R\) and \(\beta\) is the sigmoid curve slope. If the coordinates of \(n_i\) on the circle do not correspond to the image coordinates, the gray-value of this point is interpolated.

The Hessian-Affine detector was adopted. Afterwards, the regions are mapped to a circular region of constant radius and rotated in the direction of the dominant gradient, as described in [7] and [1]. Each region is divided into cells with a grid size of 4 × 4. The features are extracted for each pixel of the interest region using a CS-LMP descriptor and a histogram is built for each cell.

The parameters \(R, N\) and \(\beta\) were empirically established, and will be showed in Section IV.

As it was presented in [12], a bilinear interpolation is made over the \(x\) and \(y\) dimensions to share the weight of the feature between the four nearest cells. Thus, a 3D histogram of the CS-LMP feature locations and values is built as shown in Figure 2. The descriptor is made by concatenating the feature histograms computed for the cells, obtaining a vector of 256 positions (4 × 4 × 16). The descriptor is then normalized as in the CS-LBP presented in Section II-A.

**D. The ILBP and M-LBP methods**

ILBP proposed an improved LBP that compares all of the pixels (including the central pixel) with the mean intensity of the pixels in the patch [13].

The ILBP formulation is shown in Equation 7:

\[
ILBP_{N,R}(x,y) = \sum_{p=0}^{N-1} s(g_p - g_{mean})2^p + s(g_c - g_{mean})2^N,
\]

(7)
The Mean Local Binary Pattern (M-LBP) is similar to the ILBP, but does not consider the central pixel. In this method, the mean of a region is calculated as the threshold instead of the mean of a circular neighborhood, as shown in the Equation 9. To avoid the noise and make the feature more robust, the number of histogram bins defined for this example is 16 and thus \( b = 15 \) in Equation 9, generating codes in the range \([0, 15]\).

The construction of the descriptor was made in the same way as the CS-LMP described in Section II-C. The parameters \( R, N \) and \( \beta \) were empirically established as well for the CS-LMP as is shown in Section IV.

### IV. Experimental Evaluation

In this section, we present the performance of the M-LMP, CS-LMP and CS-LBP descriptors applied to the matching of image pairs using nearest-neighbor similarity.

The image database is available on the internet \(^1\). It has 16 sets of images with 6 different types of transformations: viewpoint change, rotation change, scale, illumination, JPEG compression and blur. The training images used in this paper are shown in Figure 5 and the test images are shown in Figure 6.

![M-LBP Diagram](image)

**Fig. 3.** M-LBP

### III. Mean Local Mapped Pattern (M-LMP)

As the original LBP is determined just by the gray level of the center pixel, some noise can be embedded in the final result. To avoid the noise and make the feature more robust, the average of the sub-region is used instead of gray level of a single pixel \(^1\). Thus, M-LMP is determined by the comparison of the pixels of a circular neighborhood to the average of this neighborhood, as shown in the Equation 9.

\[
M - LMP_{R,N,f}(x, y) = \text{round} \left( \frac{\sum_{i=0}^{N-1} f(i) P(i)}{\sum_{i=0}^{N-1} P(i)} \right) b
\]

where \( b \) is the number of histogram bins minus one, \( P(i) \) is the \( i \)th element of the weight matrix \( P \) defined in the Equation 10, and \( f(i) \) is the sigmoid mapping function defined in the Equation 11.

\[
P = [p_0p_1p_2...p_{N-1}]^T
\]

**Table I presents the evaluation of the parameter \( \beta \) for the M-LMP method applied to the image “bark”. We tested some values of the \( \beta \) parameter and we generated the recall \times precision curves, measuring its respective area under curve (AUC). For most of the test images, the best value of \( \beta \) was near 1.2 for M-LMP. For CS-LMP, the best value was 0.6 \(^1\).**

\(^1\)http://www.robots.ox.ac.uk/~vgg/research/affine/
\(^2\)http://lear.inrialpes.fr/people/mikolajczyk/
The evaluation criteria is explained in the next subsection.

### A. Evaluation criteria

The performance of the M-LMP descriptor was evaluated and compared with CS-LMP and CS-LBP, using an evaluation criterion based on Heikkilä et al. [7]. This criterion consists of the number of correct and false matches for a couple of images, knowing a priori the set of true matches (i.e., the ground truth). Heikkilä et al. [7] determine the ground truth by the overlap error of the two regions. The overlap error determines how a region A of image 1 overlaps region B of image 2. Image 2 is projected on image 1 by a homography matrix H. The overlap error is defined by the ratio of the intersection and the union of the regions given by Equation 13:

$$\epsilon_S = 1 - \frac{(A \cap H^T B) \cup (A \cup H^T B)}{255}$$  \hspace{1cm} (13)

A match is assumed to be correct if $\epsilon_S < 0.5$.

To evaluate the performance of our descriptors, we present the results with recall $\times$ precision curves [14] and their respective (AUC).

Precision represents the number of correct matches over the total number of possible matches returned (false + corrects) and is defined by Equation 14.

$$\text{precision} = \frac{n_{cc}}{n}$$  \hspace{1cm} (14)

where $n$ is the number of possible matches returned and the Euclidean distance between their descriptors is below a threshold $t_d$. The number of correct matches $n_{cc}$ includes the matched pairs whose Euclidean distance between their descriptors is below a threshold $t_d$ and $\epsilon_S < 0.5$.

Recall represents the ratio between the correct matches and the total number of true matches (the ground truth) and is defined by Equation 15.

$$\text{recall} = \frac{n_{cc}}{M}$$  \hspace{1cm} (15)

where $M$ is the number of elements in the set of true matches.

To construct the recall $\times$ precision curve, each point is defined by choosing $n \in [1, N]$, where $N$ represents the total number of possible matches. Analyzing this curve, a perfect descriptor would give a precision equal to 1 for any recall.
B. Matching results

The interest region detector used to compute the descriptors was the Hessian-Affine detector. As discussed in the previous subsection, the parameters set for the M-LMP were $\beta = 1.2$, $b = 15$, $R = 2$, and $N = 8$. For CS-LMP were $\beta = 0.6$, $b = 15$, $R = 2$, and $N = 8$, and for the CS-LBP we used the parameters $R = 2$, $N = 8$, and $T = 0.01$ [12].

We generated the recall × precision curves for all images of the test set and calculated the AUC for each of them. However, for reasons of clarity, we present only two of them in Figure 7. Note that there is a good precision for low recall for the three descriptors, i.e., the first region pairs returned are correct matches; however, as the number of region pairs returned increases, the precision decreases. We can see that the reduction of precision is more accentuated by CS-LBP. As the number of returned matches increases ($n$), M-LMP achieves a greater number of correct matches ($n_c$) compared to CS-LMP and CS-LBP.

The AUC results are shown in Table II for the M-LMP, CS-LMP and CS-LBP descriptors. Note that for all images, CS-LBP has an AUC value greater than CS-LBP and that the M-LMP descriptor has AUC values greater than the two other descriptors.

Table II. AREA UNDER CURVE (AUC)

<table>
<thead>
<tr>
<th>Image</th>
<th>M-LMP</th>
<th>CS-LMP</th>
<th>CS-LBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>east park</td>
<td>0.8801</td>
<td>0.8758</td>
<td>0.7842</td>
</tr>
<tr>
<td>laptop</td>
<td>0.9374</td>
<td>0.9219</td>
<td>0.8891</td>
</tr>
<tr>
<td>inria</td>
<td>0.8863</td>
<td>0.8798</td>
<td>0.8110</td>
</tr>
<tr>
<td>zmars</td>
<td>0.9702</td>
<td>0.9659</td>
<td>0.9499</td>
</tr>
<tr>
<td>zmonet</td>
<td>0.9609</td>
<td>0.9659</td>
<td>0.9364</td>
</tr>
<tr>
<td>resid</td>
<td>0.9090</td>
<td>0.9043</td>
<td>0.8332</td>
</tr>
<tr>
<td>leuven</td>
<td>0.9556</td>
<td>0.9479</td>
<td>0.9242</td>
</tr>
<tr>
<td>wall</td>
<td>0.9684</td>
<td>0.9639</td>
<td>0.9498</td>
</tr>
</tbody>
</table>

Table III presents the $n$ best returned correspondences by the three descriptors. For example, for the image named “resid”, taking into account the first 1027 possible matches returned, we have 583 correct matches for the M-LMP, 580 for CS-LMP and 545 for the CS-LBP descriptor. The image “east park” presents the difference of 56 correct matches favorable to the M-LMP and for the image “zmars”, our proposed approach outstands by 28 correct matches over the CS-LBP.

Figure 8 presents the correct matches achieved between the image pair called “laptop”. It was performed to check the images test results when applying the descriptors M-LMP, CS-LMP and CS-LBP.

V. CONCLUSION

In this paper, we proposed a novel method for feature description based on local pattern analysis. The method combines the CS-LBP, ILBP and M-LBP using the LMP methodology. It compares the pixels to the neighborhood average.

M-LMP features has many properties that make it well-suited for this task: a relatively short histogram and tolerance to rotation and scale change. Furthermore, it does not require setting many parameters. The performance of the M-LMP descriptor was compared to the CS-LMP and CS-LBP descriptor in terms of recall × precision curves. As M-LMP captures the gray-level nuances of the images better, its the recall × precision curves show a better performance than CS-LMP and CS-LBP methodology. We can conclude that M-LMP is a promising descriptor, with better performance over the CS-LMP and CS-LBP descriptor. Among the 10 tested images, the recall × precision curves were greater for all images for M-LMP than CS-LMP and CS-LBP. Moreover, the number of correct correspondences outperformed the CS-LMP and CS-LBP.

ACKNOWLEDGMENT

The authors would like to thank the São Paulo Research Foundation (FAPESP), grant #2011/18645-2, and National Council for Scientific and Technological Development (CNPQ) for their financial support of this research.

REFERENCES

Fig. 7. Results of the recall × precision curves for CS-LMP, M-LMP and CS-LBP descriptors.

Fig. 8. Matches achieved between the image pair called “laptop”. Total of 323 correspondences.