Wavelet Local Feature (WLF) Pattern Recognition System

Carolina Barajas-García, Selene Solorza-Calderón

Universidad Autónoma de Baja California, Facultad de Ciencias,
Tijuana-Ensenada, Baja California,
Mexico

{cbarajas,selene.solorza}@uabc.edu.mx

Abstract. In this paper is presented a pattern recognition methodology based on local feature extraction. The purpose of this system is to identify and locate, in three different scale pyramids, key points that represent relevant information of the image; this information is stored in a descriptor which is used to compares the key points of two images and know if they have similar information, or if they are the same images. This methodology uses the Haar wavelet transform to generate the three scale pyramids. This transform is used because it has several properties, such as noise elimination, multi-resolution analysis, and detection of diagonal, horizontal and vertical edges. The performance of this system was tasted using images with different scales and comparing the results with the Scale Invariant Feature Transform (SIFT) and Speeded-Up Robust Features (SURF) methodologies. The WLF system showed to has the highest percentage of correct point-matching.

Keywords: wavelets, local features, pattern recognition system, SIFT, SURF.

1 Introduction

Since the middle of last century, the extraction of features in image have been an active area in the pattern recognition field. A feature is a piece of information that is relevant for the resolution, through digital images, of some specific computational task. The features can be structures in the image like points and edges. These features are included in a descriptor, which specifies elementary properties of the object, such as shape, color, texture, among others. An example of a system based in feature extraction is SIFT (Scale Invariant Feature Transform), introduced by Lowe in 2004 [1,2].

SIFT is widely used, because of the innovative idea of using key points to represent the relevant information, allowing to work with images containing

Carolina Barajas-García, Selene Solorza-Calderón

several objects, scenarios and even fragments of objects, unlike other patterns recognition systems as Principal Component Analysis (PCA) [3], or the binary mask system based on one-dimensional signatures [4,5]. However, SIFT have several disadvantages, the computation time is considerable and has low performance with images that present few amount of noise or inhomogeneous illumination. In addition, it only allows the use of grayscale images and its generalization for the classification of color images is complicated and uses a considerable computational time.

There are variants of SIFT that try to improve its performance and their deficiencies. The most important is SURF (Speeded-Up Robust Features), presented by Herbert Bay in 2006 [6], which is a fast version of SIFT, it reduces the computational cost time but decreases its efficiency considerably. Another variant is WSIFT, proposed by Lim and collaborators in 2009 [7], based on the wavelet transform. This work shows a comparative analysis of the performance of the system using the Haar, Daubechies and Gabor wavelets transforms. The WSIFT system is not an improvement of SIFT, since it continues to maintain the same problems.

The proposal for this work is to construct a pattern recognition systems based on local features extraction using the Haar wavelet transform, since it has several properties such as excellent performance to detect and eliminate noise, 2D multiresolution analysis, scale and translation variables, that allow us to improve the system's efficiency. The rest of this work is organized as follows: Section 2 presents the Haar wavelet transform. Section 3 describes the procedure for the WLF pattern recognition system. Section 4 shows the tests and results. Finally, conclusions are given in Section 5.

2 The Haar Wavelet Transform

Given a function $f(t) \in L^2(\mathbb{R})$, that is f is an square-integrable function, the wavelet transform is defined as:

$$W_{\psi_{a,b}}\{f(t)\} = \int_{-\infty}^{\infty} f(t)\overline{\psi_{a,b}(t)}dt, \tag{1}$$

where

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), \tag{2}$$

being ψ the wavelet analysis and a, b the scaling and translational parameters, respectively.

In this work the Haar wavelet was used, which is defined as follows:

$$\psi(t) = \begin{cases} 1, & 0 \le t < \frac{1}{2}, \\ -1, & \frac{1}{2} \le t < 1, \\ 0, & \text{other way.} \end{cases} \tag{3}$$
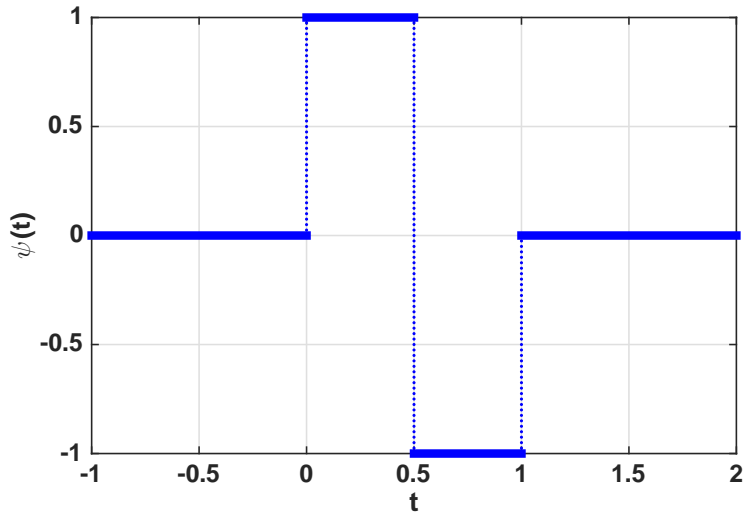
Fig. 1. Graph of the Haar wavelet transform.

The graph of the Haar wavelet transform can be seen in Fig. 1.

When the wavelet is applied in two dimensions, the image is splitted into four sub-images (that are half of the size of the given image), with different frequencies: high-high (HH), high-low (HL), low-high (LH) and low-low (LL). To apply the wavelet transform on more than one level, one of the four sub-images is selected and the same procedure is performed, this procedure is known as 2D multiresolution analysis (2D MRA). The 2D MRA procedure sketch is given in Fig.2. In this work it was used the sub-images HH, LH and HL, Fig. 3 shows that.
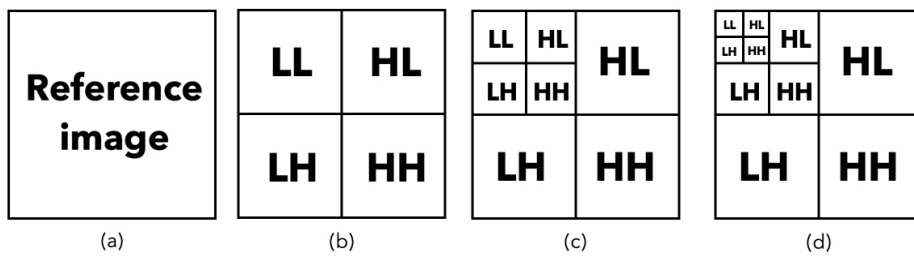

Fig. 2. Sketch of the wavelet transform levels (2D multiresolution analysis). (a) Reference image. (b) First level of the wavelet transform. (c) Second level. (d) Third level.
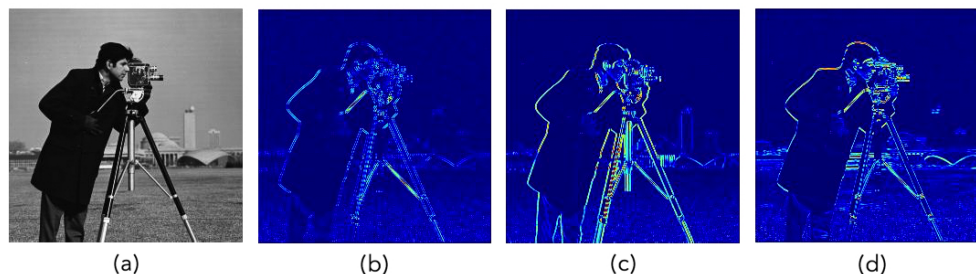
127 *Research in Computing Science* 147(4), 2018

Fig. 3. Sub-images of the Haar wavelet transform. (a) Reference image. (b) HH frequencies. (c) LH frequencies. (d) HL frequencies.

3 Wavelet Local Feature (WLF) System

This section describes the procedure to build the WLF system, which consists of three steps:

1. Generation of wavelet pyramids.
2. Localization of key points.
3. Generation of key point descriptor.

3.1 Generation of Wavelet Pyramids

The purpose of construct a wavelet pyramid is to create a space where the most important details of an image are highlighted, using the wavelet transform this is possible, since this transform enhances the edges of the image. Three wavelet pyramids will be generated, using the three sub-images obtained from the Haar wavelet transform (Fig. 3), as was done in section 2.

The procedure for generating the first pyramid, P_{HH}, consists of taking the reference image and applying the Haar wavelet transform in 5 levels using only the HH frequencies, in this way the diagonal details of the image are enhanced, Fig. 4a. The second pyramid, P_{LH}, is generated by taking only the LH frequencies, highlighting the vertical details, Fig. 4b. Finally, the third pyramid, P_{HL}, is created using the HL frequencies, enhancing the horizontal edges, Fig. 4c.

3.2 Localization of Key Points

The first step for locating the key points is filters each element into the pyramid with a binary edge mask. For example, the first level of P_{LH} is filtered using the binary edge mask M^1, that is of the same size of P^1_{LH}. Mathematically this is:

$$F^1_{LH} = P^1_{LH} \circ M^1, \qquad (4)$$

where the super index indicates the level of the pyramid and $\circ$ means an element-wise product or Hadamard product [8]. The process is shown in Fig.5.
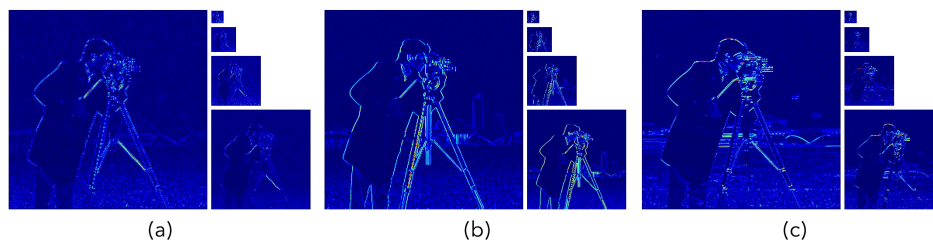
Fig. 4. Wavelet pyramids. (a) P_{HH} pyramid. (b) P_{LH} pyramid. (c) P_{HL} pyramid.

The filtering process is performed in the three pyramids like:

$$F_{HH}^n = P_{HH}^n \circ M^n, \tag{5}$$

$$F_{LH}^n = P_{LH}^n \circ M^n, \tag{6}$$

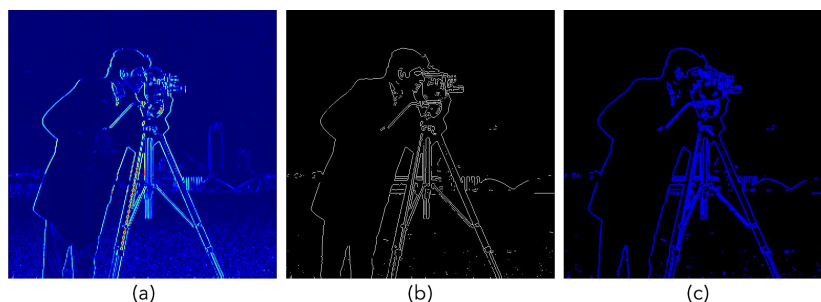$$F_{HL}^n = P_{HL}^n \circ M^n, \tag{7}$$

where $n = 1, ..., 5$.

Fig. 5. (a) P_{LH}^1. (b) Edge binary mask M^1. (c) Filtered image.

Once the three pyramids are filtered with the edge binary masks, the next step is to analyze if these points are maximum or minimum.

This is done by comparing each pixel with its 8 neighbors at the same level and the 9 neighbor's into each adjacent scales. If the selected pixel is the maximum or minimum of the 27 points, then it is taken as a key point. Figure 6 shows the key points obtained for the three wavelet pyramids.

3.3 Generation of Keypoint Descriptor

The first step to generate the descriptor is to take a neighborhood of 20×20 pixels around the key point and divide it into 16 sections of 5×5 pixels (Fig. 7a).

Carolina Barajas-García, Selene Solorza-Calderón

Fig. 6. (a) Key points detected in P_{HH}. (b) Key points detected in P_{LH}. (c) Key points detected in P_{HL}.

Then a Haar-like filters d_x and d_y (Fig. 7b and c, respectively), are applied to each of the sections:

$$fx_n = S_n \circ d_x, \tag{8}$$

$$fy_n = S_n \circ d_y, \tag{9}$$

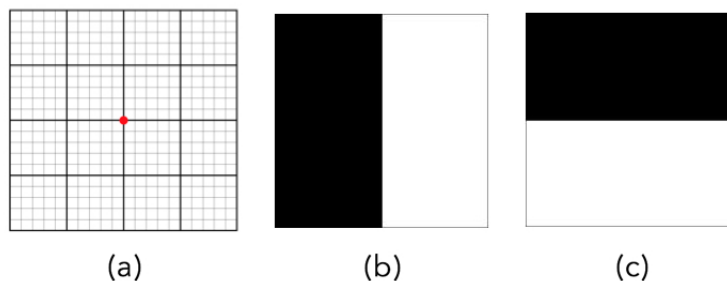where S_n is the 5×5 pixel sections and $n = 1, ..., 16$.

Fig. 7. (a) Neighborhood around a key point. (b) Haar-like filter d_x. (c) Haar-like filter d_y.

The addition of all intensity values of fx_n, fy_n, $|fx_n|$ and $|fy_n|$ are obtained, in this manner 4 scalar values are given for each of the 16 sections, having a total of 64 values. These 64 values form the key point descriptor [6].

4 Tests and Results

In order to be able to compare two images and know if they share information, the key points of the reference image and the problem image were located. Then,

the descriptors for all these key points were generated. After that, the euclidean distance of one descriptor of the reference image with all descriptors of the problem image are calculated.

The pair that has the minimum distance and also if that minimum distance satisfies to be less than a threshold of 0.15, the pair will be labeled as a match. The same procedure is applied to the rest of the descriptors of the reference image.

The WLF system was tested using the reference images of Fig. 8 with 10% and 15% reduced versions of that images. The performance of the WLF system was compared with SIFT and SURF systems, noticing that the thresholds values for the euclidean distance for SIFT and SURF are 0.8 and 0.3, respectively. Table 1 shows the results of the WLF system comparing the reference images in Fig. 8 with their reduced versions. The third row shows the number of key points that made match. In the fourth row are the number of key points that made match and are correct. In the last row the percentage of correct matched key points are given. In Table 1 is observed that for the 10% reduced image the percentage of correct matched key points is 92% and for the 15% reduced image the percentage is approximately 91%. Tables 2 and 3 show the results obtained for the SIFT and SURF systems, respectively, showing that the SIFT system presents a percentage of 89% for a 10% reduced image and 85% for a 15% reduced image. Finally, for the SURF system, 75% and 64% was obtained for the 10% and 15% reduced images, respectively.

Fig. 8. (a) Reference image: Camera. (b) Reference image: Lena.

Another result that is observed in Tables 1 to 3, is the number of key points that are detected in the different systems. The WLF system detects almost three times more points than SIFT, this is to be expected since in WLF there are three pyramids to locate key points while in SIFT there is only one. For this reason there are more key points that represents more information of the image, which leads to a higher percentage of correct matched key points. The SURF system detects few points, compared to WLF and SIFT, so the percentage of correct matched points is low. Here it was reaffirmed that by reducing the computational cost, the efficiency decrease considerably.

Table 1. WLF system.

	Camera		Lena	
	10%	15%	10%	15%
# matched key points	840	705	993	734
# correct matched key points	775	641	915	670
% correct matched key points	92.26%	90.92%	92.15%	91.28%

Table 2. SIFT system.

	Camera		Lena	
	10%	15%	10%	15%
# matched key points	255	221	289	249
# correct matched key points	229	184	259	222
% correct matched key points	89.8%	83.26%	89.62%	89.15%

Table 3. SURF system.

	Camera		Lena	
	10%	15%	10%	15%
# matched key points	171	166	151	151
# correct matched key points	136	112	109	90
% correct matched key points	79.53%	67.47%	72.18%	59.6%

5   Conclusions

The WLF pattern recognition system presented in this work shows a higher percentage of correct matched key points, which means that the system is correctly matching two images even if one of them is reduced by 10 % or even 15 %. This system uses the Haar wavelet transform for its great performance by enhancing edges and highlighting relevant image information. This wavelet transform is used to generate the wavelet pyramids which are useful for locate the key points. To compare images, a descriptor was created using Haar-like filters. The methodology proposed in this work was compared with SIFT and SURF systems, showing that the WLF is the system that presents a greater percentage of correct matched points.

References

1. Lowe, D. G.: Object recognition from local scale-invariant features. The proceedings of the seventh IEEE international conference, 2, pp. 1150–1157 (1999)
2. Lowe, D. G.: Distinctive image features from scale-invariant key points. International journal of computer vision, 60(2), pp. 91–110 (2004)
3. Jolliffe, I.: Principal component analysis. John Wiley & Sons (2002)
4. Barajas-García, C., Solorza-Calderón, S., Álvarez-Borrego, J.: Classification of fragments of objects by the Fourier masks pattern recognition system. Optics Communications, 367, pp. 335–345 (2016)

5. Ventura, A. S., Álvarez-Borrego, J., Solorza, S.: Adaptive nonlinear correlation with a binary mask invariant to rotation and scale. Optics Communications, 339, pp. 185–193 (2015)
6. Bay, H., Tuytelaars, T., Van-Gool, L.: Surf: Speeded up robust features. Computer vision?, ECCV, pp. 404–417 (2006)
7. Lim, J., Kim, Y., Paik, J.: Comparative analysis of wavelet-based scale-invariant feature extraction using different wavelet bases. Signal Processing, Image Processing and Pattern Recognition, pp. 297–303 (2009)
8. Golub, G. H., Van-Loan, C. F.: Matrix Computations. 3 (2012)