Signal Processing: Image Communication 61 (2018) 85–96

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Face recognition with Patch-based Local Walsh Transform

Meryem Uzun-Per [a,*], Muhittin Gökmen [b]

[a] Istanbul Technical University, Computer Engineering Department, Maslak, 34469, Istanbul, Turkey
[b] MEF University, Computer Engineering Department, Maslak, 34396, Istanbul, Turkey

ARTICLE INFO

Keywords:
Face recognition
Local representations
Walsh Transform

ABSTRACT

In this paper, we present a novel dense local image representation method called Local Walsh Transform (LWT) by applying the well-known Walsh Transform (WT) to each pixel of an image. The LWT decomposes an image into multiple components, and produces LWT complex images by using the symmetrical relationship between them. Cascaded LWT (CLWT) is also a dense local image representation obtained by applying the LWT again to real and imaginary parts of LWT complex images. Applying the LWT once more to real and imaginary parts of LWT complex images increases the success rate especially on low resolution images. In order to combine the advantages of sparse and dense local image representations, we present Patch-based LWT (PLWT) and Patch-based CLWT (PCLWT) by applying the LWT and CLWT, respectively, to patches extracted around landmarks of multi-scaled face images. The extracted high dimensional features of the patches are reduced through the application of the Whitened Principal Component Analysis (WPCA). Experimental results show that both the PLWT and PCLWT are robust to illumination and expression changes, occlusion and low resolution. The state-of-the-art performance is achieved on the FERET and SCface databases, and the second best unsupervised category result is achieved on the LFW database.

1. Introduction

Face recognition is automatically identifying or verifying a person from a still image or a video frame. It has been studied for more than 30 years but is still a challenging subject of computer vision. The challenges in this area largely occur due to illumination, viewpoint, facial expression, scale, and resolution variances. There is a wide range of information security, surveillance, law enforcement, and entertainment applications which use face recognition despite there being more accurate biometric indicators such as fingerprint or iris [1]. However, these indicators require the cooperation of those being observed whereas face images can be taken without the requirement for any such cooperation. This property has also increased interest in face recognition as a field of study.

Face recognition systems generally have four main stages; face detection, alignment, feature extraction, and classification. Even though the most successful face detection, alignment, and classification algorithms are used, if the feature extraction algorithm does not perform adequately, the system will not be successful. In this work, we focus on the feature extraction stage, which is probably the most challenging part of face recognition systems.

The process of feature extraction can be categorized into two classes as global and local. In global approaches, features are extracted from the entire image. Global approaches include subspace approaches such as Principal Component Analysis (PCA) [2], and Linear Discriminant Analysis (LDA) [3]; the transformation from spatial space to frequency space as in Discrete Cosine Transform (DCT) [4], and Discrete Fourier Transform (DFT) [5]; as well as moment based approaches such as Zernike Moments [6]. Recently, local approaches have been used more frequently due to their success in handling variations such as illumination and expression. Local representations are also categorized into two types as sparse and dense representations [7]. In sparse local representations such as Scale Invariant Feature Transform (SIFT) [8], points of interest are detected to be used for object detection. In dense local representations, such as Local Binary Patterns (LBP) [9] and Gabor [10], features are extracted by applying the method to each pixel of an image. Unlike supervised methods, these unsupervised local representations provide a solution when there is no information other than an image. In this work, we propose another unsupervised method based on Walsh transform (WT). To this end, we first propose a dense local representation which localizes the global WT by utilizing it for each pixel of an image and then extracting complex features from this

* Corresponding author.
E-mail addresses: meryemuzunper@gmail.com (M. Uzun-Per), gokmenm@mef.edu.tr (M. Gökmen).

https://doi.org/10.1016/j.image.2017.11.003
Received 9 June 2017; Received in revised form 18 November 2017; Accepted 19 November 2017
0923-5965/

application. Then, we apply this dense local representation, Local Walsh Transform (LWT), to patches which are created around the points of interest of multi-scaled face images, and so take also the advantage of sparse local representations.

We presented some preliminary results of local Walsh based work at the EUVIP conference [11], and the SIU national conference [12]. Here, we extend these works and explain the novel Patch-based LWT (PLWT) and Patch-based Cascaded LWT (PCLWT). The main contributions of this work are summarized below:

- LWT and CLWT are the first methods which creates complex images from WT and this is the first journal paper which covers these methods. We presented the LWT and CLWT methods in [11] and [12], however here we give a more detailed explanation of them in addition to correcting minor mistakes given in those papers.
- This is the first work which introduces PLWT and PCLWT. Although we sowed the seeds of patch based methods in [12], we improved the idea and proposed the PLWT and the PCLWT in this work.
- In [11] and [12], we showed results only on the FERET database [13], whereas in this work we have used the LFW [14] and the SCface [15] in addition to the FERET database.
- The proposed methods have been tested on several databases to show their effect both in the wild and on low resolution images. This is the first work which shows results of the proposed methods on the FERET, the LFW, and the SCface. The proposed methods surpass the state-of-the-art results for the FERET and the SCface databases, and are the second best methods in the unsupervised category of the LFW database.

The rest of the paper is organized as follows. In Section 2, we review the literature about local representations and WT. In Section 3, we present an overview of WT, and explain the proposed methods. In Section 4, the experimental results are given with a comparison to the state-of-the-art results. Finally, we conclude our work in the last section.

2. Literature survey

Local representations of images arouse interest due to their success in handling variations. Among sparse local representations, Scale Invariant Feature Transform (SIFT) constructs the scale space of an object image and then difference of Gaussian is applied to extract keypoints [8]. After determining the keypoints, their orientations are assigned and these orientations are compared to locate an object within the image. LBP is among the most frequently used dense local representations. LBP, which was originally designed for texture description by Ojala et al. [16], has been used for face recognition by Ahonen et al. [9]. In this method, a binary number is determined for every pixel of an image by thresholding the pixel with its neighbors. These binary numbers are then used to create a description image. In [9], the description image is divided into non-overlapping subregions, and after calculating the histogram of each subregion separately, weighted histograms are concatenated to create the final feature vector of the image. Many works use LBP and its variants such as Local Ternary Patterns [17], Multi-scale LBP (MLBP) [18], and LBP Network (LBPNet) [19]. 2D Gabor filters [10] which produce Gabor phases and magnitudes at several orientations and scales are frequently used in face recognition algorithms. Some of the works which use Gabor filters are Histogram of Gabor Phase Patterns (HGPP) [20], and Local Gabor XOR Patterns (LGXP) [21]. There are also some works which combines Gabor filters and LBP-like operations such as Local Gabor Binary Pattern Histogram Sequence (LGBPHS) [22], and Monogenic Binary Coding (MBC) [23].

There are also some works which use the advantages of both sparse and dense local representations. In [24], SIFT is applied to patches extracted from five scales of images by visiting every pixel of these images. In [25], overlapping multiple scales of face image patches centered around facial landmarks are extracted. Features of each patch are extracted by LBP, SIFT, Histograms of Oriented Gradient (HOG), or Gabor. In [26], features of overlapping multiple scales of face image patches are determined by LBP or SIFT. Following this, a Gaussian mixture model is built on the features before face verification. These works show that using high dimensional data by taking the advantages of both sparse and dense representations increases the success rate. In this paper, we develop another hybrid method based on Walsh Transform (WT) which take advantage of both the dense and sparse methods.

Previously, WT was considered as an alternative to DCT and DFT since it concentrates the energy of an image at the upper left corner, and is generally used in compression and coding [27–30]. Works which use WT in face recognition differ considerably from our approach. In [31] and [32], Kekre et al. apply WT to face images after resizing them to the kernel size. However, it has been stated that resizing an image to the kernel size decreases the success rate. In [33], PCA, DCT and WT are applied to whole image, to row mean of the image, to column mean of the image and to both row and column mean of the image. It is stated that WT is faster than the others. In [34], features of an image are extracted by applying WT to the image globally and then down sampling to quarter of the image. This feature extraction method is applied to the image for seven times and features are extracted for each sizes of the image. In [35], after extracting features by synthetic discriminant functions, dimensions of features are reduced by truncated WT. All of these methods use WT globally and produce lower results.

In [36], Hassan et al. apply WT to overlapping blocks of an image and then apply five different formulas to the images obtained by this application. Unfortunately, the success rate is quite low. Juefei-Xu et al. use WT locally by applying WT and LBP consecutively (WT + LBP) to each pixel of periocular region of images [37–39]. In Spartans, which is among the most recent Walsh-based work, WT + LBP are applied to the periocular regions of 3D face images produced from 2D images by a 3D generic elastic model [39]. An optimal feature subset is selected using Laplacian regularized A-optimality. Finally, they design a class-specific representation based on advanced correlation filters in kernel feature space to produce a compact representation. They also use the coupled max-pooling technique that strategically combines information from both the MACE filters and the ECPSDF filters. This local usage of WT is quite different than our approach, and produces a lower success rate.

3. Methodology

3.1. Overview of Walsh transform

Walsh Transform (WT) kernel is defined as a set of mutually orthonormal basis functions with values plus and minus ones [40]. The dot product of any two distinct rows and columns of a WT kernel is zero. Transpose, conjugate, and inverse of a WT kernel is equal to itself. For the WT kernel of order N, if we denote rows by i from 0 to $N-1$, the ith row has exactly i zero crossings [41,31,11]. WT can be applied to an image as follows:

$$\mathbf{F} = \mathbf{W}\,\mathbf{f}\,\mathbf{W} \tag{1}$$

where $\mathbf{f}$ is the image, $\mathbf{W}$ is the WT kernel and $\mathbf{F}$ is the transformed image. $\mathbf{f}$, $\mathbf{W}$, and $\mathbf{F}$ are square matrices and their sizes must be equal. Fig. 1(a) shows the application of a 4×4 WT to a 4×4 matrix symbolized by letters, and the output matrix shows how each pixel is constructed. Fig. 1(b) visualizes filters for a 4×4 WT kernel which constructs a transformed image in which the white squares symbolize plus and dark squares symbolize minus.

Alternatively, WT can be applied to an image directly without creating a WT kernel [27]. By this method, each pixel of the transformed image is calculated separately by the following formulation. Thus