CS 579 Term Project Final Report

Yusuf Sait Canbaz, 21201337

January 2020

1 Introduction

In this project I will analyze the paper "A study on wrist identification for forensic investigation" [1]. My main purpose is to produce an python implementation of the proposed method in the paper. The authors of the paper provides their implementation for the paper. [1] They are using matlab. I want to make an implementation in python because the language is pretty clean and very similar to pseudo code. I think, it is handy for clear depiction of algorithms.

With the advancement of bio-metric systems, there are many modalities for identification of people. Some of the most popular and widely used bio-metric modalities are face, iris, retina, voice, keystroke, hand geometry, signature, gait etc... Since these modalities are also known by the criminals, they tend to hide their physiological/behavioral traits [1]. To make forensic investigations more easier and available, there are some studies which tries benefit from new bio-metric modalities. ECG, [3] EEG, lip-print [4], mouse dynamics, tongue prints, palm print, palm-vein, androgenic hair, skin marks are some of the proposed modalities.

2 NTU wrist image database

Originally database is only consists of original wrist images. Totally, 505 subjects, 731 wrists and 3945 images. The number of subjects is not really important here because 2 wrists of the same subject are independent. So, the number of classes depends on number of unique wrists.

[1] https://github.com/matkowski-voy/Wrist-Identification-for-Forensic-Investigation

Provided database also contains segmented images and segmented & aligned wrist images. Segmented images are result of applying segmentation algorithm to the original images. Segmented & aligned images are result of applying region of interest (ROI) extraction algorithms to the segmented images. There are 2 different ROI extraction algorithms. At the end, they use segmented & aligned images to make identification.

To use as gallery and probe sets, there are 5 sets of images: SET1, SET2, SET3, SET4 and SET5. Inside the segmented & aligned images, there are another 5 sets of images named: SET1p, SET2p, SET3p, SET4p and SET5p. These sets are result of applying the second ROI extraction algorithm.

The method consists of 4 parts. Segmentation, region of interest (ROI) extraction, feature extraction and matching. I almost finished implementation of segmentation part.

3 Algorithms

The project consists of 4 parts: segmentation, ROI extraction, feature extraction and matching. Segmentation part generates segmented images from original images. ROI extraction part generates segmented & aligned images from segmented images. Feature extraction part extracts features from segmented & aligned images. Matching part generates classifiers from extracted features and builds *recognition systems* from them.

3.1 Segmentation

Segmentation part starts with implementation of *Simple Iterative Linear Clustering* (SLIC) algorithm [2]. This algorithm generates clusters of pixels. Each cluster is as homogeneous as possible. A cluster of pixels is called as *super-pixel* or as a *segment*. Implementation of SLIC is already available in *scikit-image* python library I used that. [2] But *scikit-image* implementation does not return adjacency matrix of super-pixels. So I implemented an algorithm to get adjacency matrix for super-pixels.

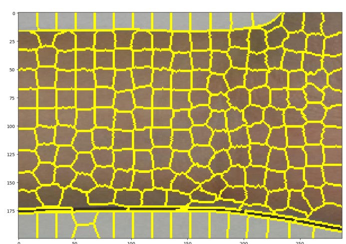

Figure 1: An image with 191 super-pixels

My algorithm *get_adjacency_matrix*, iterates over pixel pairs. A pixel pair consists of either horizontally or vertically neighboring 2 pixels. The algorithm considers heterogeneous pixel pairs. If 2 pixels of a pair belongs to different super-pixels, the pair is heterogeneous and it means the super-pixels are adjacent. To determine if 2 super-pixels are adjacent, the algorithms keeps an adjacency matrix for super-pixels. If there is m super-pixels, the matrix will be $m \times m$. $m[i,j]$ corresponds to the number of heterogeneous pairs between super-pixel i and j. A threshold on this number can be used to determine if 2 super-pixels are adjacent. Setting this threshold to a number like 5 instead of 1 makes improvement on classification.

Segmentation uses super-pixels to divide skin pixels from non-skin pixels. Super-pixels are classified as skin and non-skin. Segmentation uses some color spaces and gradient maps for feature extraction. Mean and variance of the values in color spaces and gradient maps are used as features. I only implemented *difference of Gaussian's* gradient map. I converted images to color spaces and get other gradient maps by using *scikit-image* library. To run a classification model, each super-pixel is represented by extracted features from its 8 neighbors and itself. If there is no corresponding neighbor, super-pixel's own features are used.

There are some issues with segmentation. We assume that a super-pixel is either a skin or non-skin. This totally depends on SLIC. To make adhere to skin/non-skin boundaries, high number of super-pixels are more favorable. In the segmentation, a super-pixel is labeled as skin if at least half of its pixels are skin. At the end, I get **96%** cross validation (10 fold) accuracy with **7%** standard deviation by training 100 wrist images.

3.2 Region of Interest Extraction

This part generates aligned images by using segmented images. They utilize the nature of wrist, it usually contains 2 prominent wrinkles. They try to locate wrinkles to detect the ROI. Roughly, they apply 2 different algorithms to align images. One of them assumes that 2 wrinkles lies starts and ends with upper and lower boundary of the wrist. the other does not assume such constraint. I didn't implement this part. Instead, I directly used generated segmented &aligned images.

3.3 Feature Extraction

This part extracts features from segmented & aligned images. Features are extracted by using local binary patterns (LBP), gabor filters and scale-invariant feature transform (SIFT) features. Respectively 13074, 2112 and 1280 features are extracted using LBP, gabor filters and SIFT. I couldn't extract SIFT features because I couldn't use the library used in the paper. This could reduce accuracy because they are the second most useful features [1].

LBP features are extracted from 7 different grids. Structure of grids are 7x5, 5x7, 5x5, 4x3, 3x4, 3x3 and 2x2 respectively. If there are too many cells, for example in the first 2 structure there are 35

[2] https://scikit-image.org/

cells, cell area will be small. To get more balanced LBP histograms, 2 different type of LBP features are extracted. One of them is *rotation invariant(ri)*. The other one is both *rotation invariant* and *uniform*(*riu2*). *riu2* is applied to first 2 grids since cell areas of these cell are small. *ri* is applied to the remaining 5. Then, gabor filter is applied in 16 different angles for each cell of the grids. So for each cell we get a histogram of size 16. To make calculations simpler, features and labels are saved as file.

3.4 Matching

In this part, I build classifiers and do matching with the classifiers. I build one classifier for each class. So all the classifiers are binary classifiers. They are very simple. They have the form

$$y = \beta \cdot x_n + b$$

Here β is a vector of length features. x_n is normalized features for a sample. b is bias term. If $y = 1$ means a match, $y = -1$ means a non-match. There are 2 types of classifiers, partial least squares (PLS) and linear support vector machine (SVM). To make matching easier, after I build the classifiers, I save their parameters β and b.

To match all the samples in the probe set to the gallery set, a set of classifiers are being used. So for convention, the set of classifiers will be called as *recognition system.* Different recognition systems can be built be using different representation of samples or different classifiers.s

As mentioned before, since there are 2 different ROI extraction algorithms, a unique wrist can be represented with 2 ways. So we can build 2 recognition systems one from images in SET1 and the other from images in SET1p. Both recognition systems are actually making classification on the same data. At the end, I can have 4 recognition systems. For each representation SET1 and SET1p I can built 2 types of classifiers.

4 Results

Firstly, for all the experiments I used SET1 as gallery and SET2 as probe. I also used SET1p and SET2p which are simply alternated representations of the same wrists. First, I built an SVM classifier. Critical parameter during training SVM is iteration count. With 1000 (default) iterations some classifiers might not converge. I set iteration count to To see the effect of iteration count, another set of SVMs which are trained for 10000 iterations. In addition, since there is big class imbalance, I tried to give classes weights according to their counts. Figure 2 shows performances of these SVMs. Overall SVM performances are not changed a lot.

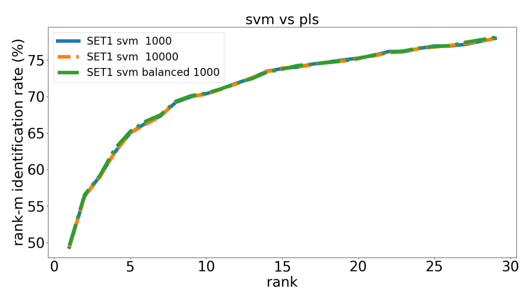

Figure 2: Cumulative Match Characteristic (CMC) curve for SVMs trained for 1000 iterations, 10000 iterations and 1000 iterations with balanced class weights

Figure 3 shows 4 possible recognition systems for a data set. It seems like overall PLS is better. Interestingly, in the paper [1] SVM performs better than PLS. This might be result of differences in SVM. Overall, SET1 performs better then SET1p. This could also observed in the paper on the results of experiment 1 (using SET1 as gallery, SET2 as probe).

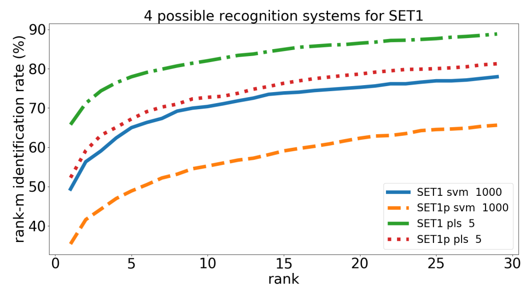

Figure 3: CMC curve for SVMs and PLSs

Figure 4 shows what happens if we use only a subset of the features using SVM classifiers. There are LBP features and Gabon features. LBP features look like very effective.

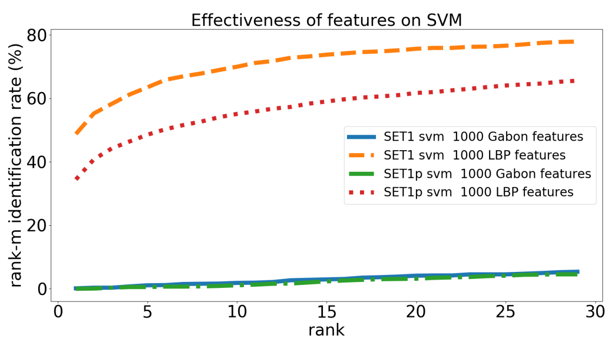

Figure 4: Effectiveness of features with SVM classifier

Figure 5 shows what happens if we use only a subset of the features using PLS classifiers. There are LBP features and Gabon features. Again LBP features look more effective.

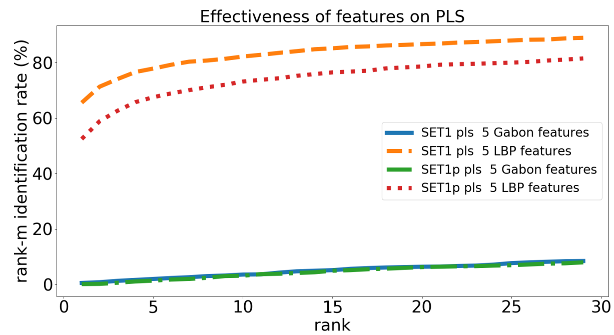

Figure 5: Effectiveness of features with PLS classifier

In figures 4 and 5, the least effective features are Gabor features. Although the paper states that least important features are Gabor features, they nearly have zero effectiveness. This is a bit suspicious.

Figure 6 shows comparison of Wrist-Meta-Match (WMM) and the best recognition system (PLS on set1). WMM utilizes all the four available recognition systems. It utilizes Extreme Value Theorem (EVT). It assumes highest value is an anomaly and fits a Weibull distribution to the remaining top n highest values. n is equal to a proportion of probe images. I used 0.2. Then it gets cumulative distribution function value for the highest value. It compares results of all 4 RSs and selects the highest.

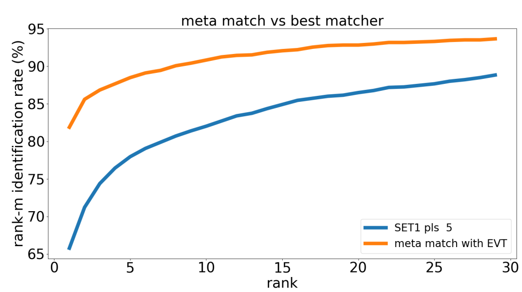

Figure 6: Performance of Wrist-Meta-Match

As you can see from figure 6, WMM is clearly more accurate. Rank-1 accuracy is 82%.

5 Conclusion and Feature Work

At the matching step, I generated 4 different recognition systems. The paper generates and ensemble of recognition systems. They call it Wrist Meta-Matching (WMM) system. Such kind of system can obviously perform better than a single recognition system.

Tuning SVM classifiers might increase the performance because SVM is more capable in the results of the paper.

Linear combination of recognition systems could also be an alternative to the ensemble. Different kind of ensemble techniques might be applied.

More elaborate feature extraction or extracting more features with LBP might help. Customized grid structures which are some how interpreted from data might be used.

Lastly, lack of SIFT features obviously decreased the performance of recognition system. Adding some other features might also help. In the paper, it states PCA kind of feature reduction techniques didn't help. So I didn't try them.

References

[1] Matkowski, Wojciech Michal, Frodo Kin Sun Chan, and Adams Wai Kin Kong. "A Study on Wrist Identification for Forensic Investigation." Image and Vision Computing (2019).

[2] Achanta, Radhakrishna, et al. "SLIC superpixels compared to state-of-the-art superpixel methods." IEEE transactions on pattern analysis and machine intelligence 34.11 (2012): 2274-2282.

[3] T.W. Shen, "Biometric identity verification based on electrocardiogram(ECG)",PhD Dissertation,University of Wisconsin, Madison, 2005

[4] Chauhan, Sushil, A. S. Arora, and Amit Kaul. "A survey of emerging biometric modalities." Procedia Computer Science 2 (2010): 213-218.

[5] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Susstrunk. "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods" PAMI. Vol 34 No 11. November 2012. pp 2274-2281.