早稲田大学大学院情報生産システム研究科

博　士　論　文　概　要

論　文　題　目

Study on Multi-Scans based Invariant Feature Descriptors and Their Applications

申　請　者

Ｗｅｉ　ＺＨＯＵ

情報生産システム工学専攻
イメージメディア研究

2013 年　02 月

Over the last several decades, pattern recognition research has attracted substantial attention and contributed to a skyrocketing growth in many applications including biometrics recognition, texture classification, etc. Generally, pattern recognition research consists of two key steps: feature extraction and matching. Some measurements of the image that form a feature descriptor are calculated in the first step. Then the image is matched based on its feature descriptor. Recently, many feature descriptors, such as Scale Invariant Feature Transform (SIFT) and Local Binary Patterns (LBP), have been developed and some major advantages have been achieved. As a result, the performance of these feature descriptors under controlled conditions has now reached a practical level. However, these feature descriptors face some challenges in real environments with variations in scale, rotation, illumination, outliers, occlusion, etc. For example, SIFT and LBP are very sensitive in the changes of non-monotonic illumination, especially under the challenges of outliers or occlusion issues where these feature descriptors cannot obtain discriminative information from the image for robust recognition.

Since the performance of a feature descriptor is directly related to the amount of variations observed from images. A feature descriptor can eliminate the influence of these variations more effectively; it can yield better recognition results. Therefore, the objective of this research aims at designing the invariant and discriminative feature descriptors under various applications. In order to deal with the challenges in scale, rotation, illumination, outliers and occlusion problems, a new multi-scans encoding model and some new projection strategy are proposed in this thesis. The multi-scans encoding model calculates the edge responses in multiple resolutions and multiple orientations at each pixel position in the image by multi-scans, generating a code from the relative strength magnitude by some encoding rule. And then a scale, rotation and illumination invariant feature descriptor can be obtained by using the direction of the most prominent edge response in related resolution space. In order to further improve the performance, the multi-scans encoding model is extended into Curvelet frequency space to better capture the curve and edge information of images, offering another invariant feature descriptor with salient visual properties such as spatial frequency, spatial localization and orientation selectivity. Meanwhile, our new projection strategy can characterize the within-class compactness and between-class separability under Fisher's linear discriminant function by L1-norm and Correntropy optimization technique when outliers or occlusion occur.

This thesis consists of 6 chapters that are briefly summarized as follows:
In Chapter 1, research background and motivation of this work are introduced, and also the principal contributions and the organization of this research are described.

In Chapter 2, based on the multi-scans encoding model, a novel invariant feature descriptor called 1D Local Patterns by Multi-Scans (1DLPMS) is proposed. Briefly speaking, 1DLPMS consists of five steps. Firstly, the gray values of pixels in the image are represented as a block. Secondly, multi-scans, which are used to capture different spatial information on the block, are applied to transform the block into several 1D sequences at multiple resolutions and multiple orientations. Thirdly, the edge responses of these 1D sequences are computed. And in the fourth step, the related 1D sequence that has the largest edge response is encoded into one code by some encoding rule, which is invariant to scale, rotation and illumination variations besides preserving the appearance of images embedded in the original gray scale. At last, by considering all the blocks in the image, a histogram can be obtained as the feature descriptor of this image where horizontal coordinate is the index of codes and vertical coordinate is the magnitude of the codes. In the experiments, based on a general database including buildings, scenes, faces, textures and so on, the results demonstrate that our proposed 1DLPMS outperforms the conventional feature descriptors, such as SIFT and LBP, under various challenging situations.

In Chapter 3, since the above proposed 1DLPMS is focus on the local information, which is not good at capturing the global feature of an image. And the codes used in 1DLPMS are predefined, which are not suitable in any pattern recognition tasks. Thus, in this chapter, the multi-scans encoding model is extended into Curvelet frequency space. The Curvelet transformation and multi-scans encoding model are mutually complementary because multi-scans encoding model can capture the details of local appearance, whereas Curvelet transformation extracts global information from the image. Based on Curvelet transformation and multi-scans encoding model, one novel invariant local Curvelet feature descriptor is proposed. This feature descriptor not only considers the global information of image, but also designs several pattern-specific codes from Curvelet filtered images. For the sake of solving outliers or occlusion issues as well as selecting compact and discriminative features, the above invariant local Curvelet feature descriptor is further processed by our proposed Linear Discriminant Analysis based on L1-norm (LDA-L1) method. The conventional linear discriminant analysis approach, which is based on L2-norm, is sensitive to outliers or occlusion problems. It also often suffers from the so-called small sample size (3S) problem. Compared to conventional L2-norm based linear discriminant analysis method, our proposed LDA-L1 has several advantages: 1) It is robust to outliers or occlusion. 2) It has no 3S problem. 3) It is invariant to rotation. 4) It takes less time complexity. Finally, performance assessments in several databases under different challenges show the effectiveness of the extended invariant feature descriptors.

In Chapter 4, face recognition problems are evaluated by our proposed invariant feature descriptors. One novel and effective matching method called Weighted Histogram Spatially constrained Earth Mover's Distance (WHSEMD) is proposed. The proposed matching method partitions the source facial image into non-overlapping parts while represents the destination facial image as a set of overlapping parts at different positions. One part in the source facial image can be matched only to one of its neighboring parts in the destination facial image under a spatial constraint. Thus, WHSEMD can solve small pose issue in the face recognition problems. WHSEMD also takes the discriminative powers of different facial parts into consideration. The features of important parts in the face, such as eyes, nose and mouth, give higher impact to the final matching score than others. Based on our proposed invariant feature descriptors and matching method, huge of experiments including face identification, gender estimation and facial expression recognition problems are estimated. From these experiments, the effectiveness and power of the proposed approaches are demonstrated, especially under the challenge of facial occlusion problem. In general, the recognition rate can be improved about 3%-5% compared to the state-of-the-art face recognition methods.

In Chapter 5, texture classification based on our proposed invariant feature descriptors is estimated. If the proportion of outliers is more than 40% in the image, the above-proposed LDA-L1 approach still cannot solve this issue effectively. Thus, in order to solve large outliers problem, another novel linear discriminant analysis based on a new Maximum Correntropy Criterion (MCC) optimization technique is proposed. In information theory learning, it has been pointed out that L2-norm and L1-norm are the global measurements while MCC is a local one. By global, it means that all the data points in the joint space will contribute equally to the value of the measurement, and the locality of MCC means that the value is mainly determined by the kernel function. Since an outlier is far away from the data cluster, then its contribution to estimate correntropy will be smaller so that it always receives a low value in the measurement. Therefore, the outliers will have weaker influence on the estimation as correntropy increases. Thus, MCC based linear discriminant analysis is capable of analyzing non-Gaussian noise to reduce the influence of large outliers substantially while keeping most discriminative information. In the evaluations, our proposed approaches show good performance with various variations, especially under the challenge of large outliers in the application of texture classification. Compared to some famous approaches, the accuracy of classification can be improved about 6% and 1% on Brodatz and CUReT databases, respectively.

In Chapter 6, this thesis is concluded and some future works, such as extending our proposed methods to kernel or 3D space, are discussed.