Feature Selection on Affine Moment Invariants in Relation to Known Dependencies

Aleš Zita[1,2], Jan Flusser[1], Tomáš Suk[1], and Jan Kotera[1,2]

[1] Institute of Information Theory and Automation*, Czech Academy of Sciences, Prague, Czech Republic
zita@utia.cas.cz
[2] Charles University in Prague**, Faculty of Mathematics and Physics, Czech Republic

Abstract. Moment invariants are one of the techniques of feature extraction frequently used for pattern recognition algorithms. A moment is a projection of function into polynomial basis and an invariant is a function returning the same value for an input with and without particular class of degradation. Several techniques of moment invariant creation exist often generating over-complete set of invariants. Dependencies in these sets are commonly in a form of complicated polynomials, furthermore they can contain dependencies of higher orders. These theoretical dependencies are valid in the continuous domain but it is well known that in discrete cases are often invalidated by discretization. Therefore, it would be feasible to begin classification with such an over-complete set and adaptively find the pseudo-independent set of invariants by the means of feature selection techniques. This study focuses on testing of the influence of theoretical invariant dependencies in discrete pattern recognition applications.

Keywords: Affine invariants, Image moments, Feature selection, Machine learning, Pattern recognition

1 Introduction

One of the difficult tasks in image processing is a recognition of shapes degraded by some transformation. Several approaches to the invariant recognition exist. Those are methods based on brute force, i.e. methods which are focused on training data alteration in such way that they add artificial training data deformed with all possible transformations in question. This approach is adopted for example by deep convolutional networks. It has several disadvantages such as impossibility to generate all of the possible transformations or to cover some of the transformation classes. Approach that overcomes these disadvantages is

* This work was supported by Czech Science Foundation (GA ČR) under grant GA15-16928S.
** This work was supported by The Charles University Grant Agency (GA UK) under grant 1094216.

using computed features which are mathematically invariant to a certain family of transformations.

General moment is defined as a projection of a function to polynomial basis

$$M_{pq} = \iint P_{pq}(x,y) f(x,y) \mathrm{d}x \mathrm{d}y.$$

Moment invariant is then a function of moments satisfying invariance to particular class of deformations. As a simple example of invariant's creation we can demonstrate construction of invariants to translation from geometric moments. We define geometric moment as

$$m_{pq} = \iint x^p y^q f(x,y) \mathrm{d}x \mathrm{d}y,$$

with only m_{00} being invariant to translation. We can construct other invariants using m_{00} as follows

$$\mu_{pq} = \iint (x - x_t)^p (y - y_t)^q f(x,y) \mathrm{d}x \mathrm{d}y,$$

where $x_t = m_{10}/m_{00}$ and $y_t = m_{01}/m_{00}$.

This technique is equivalent to shifting the center of gravity to the origin. Moments μ_{pq} are called central moments. Similar approaches can be used to create invariants to rigid or affine transform. For more examples please refer to [1–3]. Several techniques of affine moment invariants generation exist. Most of them produce over-complete sets containing dependencies.

In our experiments, we use the graph method of generating affine invariants. The core of this approach is in generating invariants using undirected multigraphs [3].

If we define 'cross product' of two image points (x_1, y_1) and (x_2, y_2) as

$$C_{12} = x_1 y_2 - x_2 y_1,$$

then after the image affine transform it holds

$$C'_{12} = J \cdot C_{12},$$

where J denotes the Jacobian. Therefore, C_{12} is relative affine invariant. Basic idea of invariants creation is integration of the cross product as a moment. After Jacobian elimination by normalization we get the affine invariants. This means that for $N \geq 2$ (degree of the invariant – the number of moments multiplied in one term), we can define

$$I(f) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \prod_{k,j=1}^{N} C_{k,j}^{n_{kj}} \cdot \prod_{i=1}^{N} f(x_i, y_i) \mathrm{d}x_i \mathrm{d}y_i, \tag{1}$$

where $n_{k,j}$ are nonnegative integers. After affine transform, we get

$$I' = J^w |J|^N \cdot I, \tag{2}$$

where $w = \sum_{jk}^{N} n_{jk}$ is the invariant's weight. Normalizing by μ_{00}^{w+N} gives an invariant

$$\left(\frac{I}{\mu_{00}^{w+N}}\right)' = (\mathrm{sign}(J))^w \left(\frac{I}{\mu_{00}^{w+N}}\right). \tag{3}$$

For example $N = 2; n_{12} = 2$ gives

$$I(f) = \int\limits_{-\infty}^{+\infty} \cdots \int\limits_{-\infty}^{+\infty} (x_1 y_2 - x_2 y_1)^2 f(x_1, y_1) f(x_2, y_2) \mathrm{d}x_1 \mathrm{d}y_1 \mathrm{d}x_2 \mathrm{d}y_2 =$$

$$= 2(m_{20} m_{02} - m_{11}^2).$$

Every affine invariant created by this method can be represented by a multigraph where each point (x_k, y_k) corresponds to a vertex, and the cross product $C_{jk}^{n_{jk}}$ corresponds to a multiedge with multiplication factor n_{jk} connecting vertices k and j. Then the problem of affine invariant generation is equivalent to problem of generation single connected component multigraphs with w edges and number of vertices grater or equal to 2.

1.1 Dependencies and completeness

In the present time, the search for the dependencies among moment affine invariants generated by the graph method is based on brute force approaches. First, the duplicities in generated invariants are found by reduction of polynomials to their irreducible forms. Next, the trivial dependencies are eliminated (zero invariants), after that the linear and polynomial dependencies are searched using brute force algorithm. This is a time costly process, because the whole invariant space needs to be searched. Therefore, finding higher order dependencies is practically impossible.

For example, from all 2 533 942 752 generated invariants of order ≤ 12 there are 2 532 349 394 zero invariants and 1 575 126 identical invariants, 14 538 linear combinations and 2 105 products. After this first reduction, there are still 1 589 irreducible invariants from which we know that only 80 are independent.

The cardinality of complete and independent invariant set can be calculated using following formula [4]

$$c = \binom{r+d}{r} - DOF(T), \tag{4}$$

where r denotes the order of invariant; d is number of image dimensions and $DOF(T)$ denotes the degrees of freedom of the degradation operator T (in our case the affine transform). In this way we can easily calculate, that for moment invariants of 4th order and affine transform of 2D image the complete independent set has cardinality equal to 9. In other words, there are independent sets of cardinality 9 within $I_1 - I_{32}$ (affine invariants of order 4 generated by the graph method). Some of these sets are known and had been proven to form a complete set, e.g. $\{I_1, I_2, I_3, I_4, I_6, I_7, I_8, I_9, I_{22}\}$ [5].

In our study, we want to investigate the relations between theoretical properties of affine invariants in combination with practical methods of feature selection. Therefore, we designed our experiments to test the strength of known dependencies against discriminative powers of individual invariants.

From the known set of dependencies of 4th order, 5 can be produced by set of 9 invariants. Let d1 - d5 denote the following known dependencies (corresponding to dependencies no. 1 ,2 ,6 ,8 and 12 in [5]):

$$
\begin{aligned}
&d1: -4I_1^3I_2^2 + 12I_1^2I_2I_3^2 - 12I_1I_3^4 - I_2I_4^2 + 4I_3^3I_4 - I_5^2 = 0\\
&d2: -16I_1^3I_7^2 - 8I_1^2I_6I_7I_8 - I_1I_6^2I_8^2 + 4I_1I_6I_9^2\\
&\qquad + 12I_1I_7I_8I_9 + I_6I_8^2I_9 - I_7I_8^3 - 4I_9^3 - I_{10}^2 = 0\\
&d3: -4I_1I_2I_9 + 4I_1I_{16}^2 + I_2I_8^2 + 4I_3^2I_9 - 4I_3I_8I_{16} + I_{18}^2 = 0\\
&d4: -I_1I_2I_{15} - I_1I_2I_{16} + 2I_1I_3I_11 + I_2I_{22}\\
&\qquad + I_3^2I_{15} + I_3^2I_{16} - 2I_3I_{32} - I_4I_{11} = 0\\
&d5: 2I_1I_3I_{24} + I_1I_{15}I_{17} - I_4I_{24} - I_{15}I_{28} - I_{17}I_{22} + I_{18}I_{22} = 0
\end{aligned}
\qquad (5)
$$

2 Method

We proposed several experiments to test whether the theoretical relations between individual image affine moment invariants are reflected in discrete world of pattern recognition tasks. For this purpose, we utilize well known feature selection algorithms, classifiers and datasets.

2.1 Feature Selection Algorithms

Sequential Forward Selection (SFS) A method which starts with an empty set and then sequentially selects features with best possible classification outcome. It is a basic method of selecting relatively good subset of features for low dimensional problems. Its main advantage is its computational speed. But the most prominent disadvantage is that the algorithm does not allow to remove any feature previously selected.

Sequential Forward Floating Search (SFFS) An algorithm, which in each turn adds the most significant feature and then repeatedly tries to remove features by comparing the performance with the best performance achieved so far for the same-sized subset. This way, it tries to deal with fore-mentioned disadvantage of greedy approach of SFS [6].

2.2 Classifiers

Support Vector Machine (SVM) Support Vector Machine [7] is one of the widely used and well performing classification algorithm. In our experiments, we used SVM with RBF kernel. Parameters which were used were tuned to give best possible classification performance on full feature set for given problems.

Neural Network (NNET) As a second classification algorithm, we used fully connected classification neural network with two hidden layers with 50 neurons both. The network was finely tuned to give best classification performance on both the problems. The reason behind using neural network for this task is its theoretical ability to discover complex dependencies.

2.3 Datasets

MNIST A well-known database with handwritten digits [8]. The dataset consists of 60 000 training and 10 000 testing digits images (see Figure 2.3 for illustration). For the purpose of this study, we calculated first 32 normalized affine moment invariants for all images in the dataset.

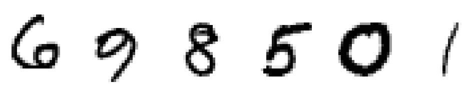

Fig. 1. Examples of MNIST database with handwritten digits. White corresponds to zero and black to one.

MEW 2014 The next database we used for our experiments is a database of segmented tree leaves [9, 10]. The affine moment invariants were calculated on the segmented images directly (see Figure 2.3 for illustration). The database contains 15 074 images distributed to 201 classes. In each of the experiments we used subset of 100 classes to reduce computational time and complexity of the classification task.

3 Experiments

3.1 SFS

The experiments were designed to investigate probability with which the feature selection algorithm selects the set of invariants from all 4th order invariants that does not produce any of the known polynomial dependency. There is a known set of 32 irreducible affine invariants and the set of known dependencies [5]. All invariants were calculated from the original binary images from MNIST

Fig. 2. Examples of MEW 2014 database with segmented tree leaves (white=0, black=1).

and MEW 2014 databases. Because of the magnitude differences of individual invariants, it is usual to normalize them prior to classification. The normalization technique used in this work is based on two phase technique

$$\mu'_{pq} = \mu_{pq} \cdot \pi^{\frac{p+q}{2}} \cdot \left(\frac{p+q}{2} + 1\right) \tag{6}$$

$$I' = \mathrm{sign}(I) \cdot \sqrt[d]{|I|}, \tag{7}$$

where d denotes degree of the invariant. First, the moments within each invariant are normalized in such a way that their corresponding complex moments are equal to 1 when calculated on unitary circle (6). Next, the invariants are normalized to degree (7). This covers the cases in which the products of many moments within a invariant can result in very large numbers. Furthermore, each invariant was scaled by a learned factor to produce the best classification performance on both the databases independently.

We started with the SFS method to progressively select most discriminative feature (invariant), one at a time. The feature selection process was forced to continue after peak classification performance was reached, until set of 9 features was selected. To introduce diversity to the experiment, each feature selection process was executed on random subset of classes.

Subsequently, we performed a search for defined dependencies (Section 1.1) on the resulting sets. Our goal was to estimate the influence of theoretical dependencies of the invariants in continuous domain to discrete world of machine learning. The statistics of feature selection process can be viewed as the indicators of discriminative powers of individual moment invariants.

3.2 SFFS

Our next effort was to improve the feature selection performance by using SFFS method, again on both datasets. But, in this case we omitted the usage of the NNET classifier. The reason being, that the outcome of neural network classifiers depends on random initialization and the optimization function of neural network has typically many local minimums, and cannot provide the level of classification consistency SFFS process requires to run efficiently. To successfully utilize neural networks for SFFS would mean to run the classification many times over to produce meaningful statistics of current classification performance. This would be impractical and time consuming.

The outputs of the experiments are the same statistics as in case of SFS.

3.3 Adding dependent feature

Our next task was focused on studying the strength of particular invariants discriminability vs. invariant dependency. We took the histogram of all selected features in previous experiments as a measure of each invariants discriminative power. Furthermore, we performed uncorrelated estimation of each invariants discriminative power by running classification statistics on sets represented by single invariants at a time.

For all known dependencies of invariants of order 4, we started the feature selection process with all the invariants from given dependency except the one with highest discriminability. This experiments goal is to study the level of particular dependency when in contradiction to strong discriminative ability.

4 Results

4.1 SFS

Starting from empty sets we have run 200 sequential feature selection trials on MNIST datasets with both classifiers. One of these runs (what is 0.5% cases) ended up with invariant set producing dependency d1, see Eq. (5). Histogram of selected invariants can be seen in Figure 3 left.

The second part of this experiment was to run the same task for MEW 2014 database. The dependent set was again generated in 1 case out of 200 (0.5%) with dependency d2 in Eq. (5). See Figure 3 right for resulting feature histogram.

Note that the sets of selected invariants differ, because discriminative powers of individual invariants changes with the classification task.

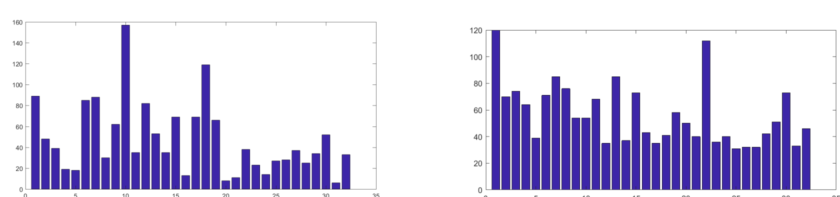

Fig. 3. Histogram of affine invariants selected by the SFS process. The images indicate relative discriminative powers of invariants $I_1 - I_{32}$(x-axis) when used in classification task of MNIST (left) and MEW 2014 (right) datasets.

4.2 SFFS

In first batch of all 200 feature selections on MNIST database, one (0.5%) resulted in dependent set being generated. In this case dependency d2 in Eq. (5) emerged.

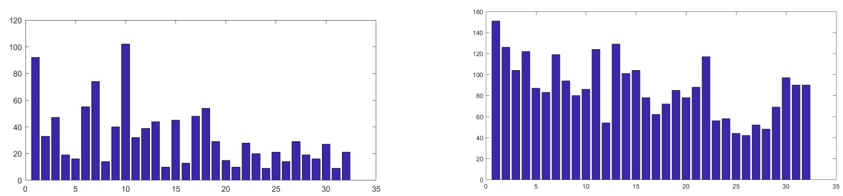

Fig. 4. Histogram of affine invariants selected by the SFFS process. The images indicate relative discriminative powers of invariants $I_1 - I_{32}$(x-axis) when used in classification task of MNIST (left) and MEW 2014 (right) datasets.

The next experiment was the same configuration run on MEW 2014, resulting in 4 dependent sets being generated out of 200 trial runs, all of them having dependency d2 in (5). Histogram of both experiments can be seen in Figure 4, resp.

4.3 Adding dependent feature

Because the feature selection processes in our experiments produced only dependencies d1 and d2, see Eq. (5), we will focus in this experiment on generating those two. We estimated discriminability for the individual affine invariant by running classification on each of them separately. Because the discriminative strength of invariants is data-related, we performed the calculation for each dataset independently. We assumed, that the most discriminative invariant overall for both the dependencies is the invariant I_1, as it represents image reference ellipse and due to relatively small polynomial exponents is producing smallest numerical computational instabilities.

Our experiments showed that invariants with greatest discriminative power are I_{18} and I_{22} respectively. However, when we study the generation of dependency d1 in Eq. (5), we found I_4 to be most discriminative within the dependent group (see Figure 5 top). For the study of dependency d2 in Eq. (5), we found the I_1 to be the one with greatest discriminative power (see Figure 6 top).

Dependency d1 We ran SFS process initialized not with an empty set, but with set of $\{I_1,\ I_2,\ I_3,\ I_5\}$ (i.e. removing the strongest I_4 from dependent set) for both dataset and both the classifiers to test the strength of the dependency d1 in (5).

Fig. 5. Statistics on invariants discriminative powers when invoking d1 (5) dependency. Left: MNIST dataset, right: MEW 2014; top: the mean classification accuracies for individual invariants $I_1 - I_{32}$; middle top: the mean classification accuracies for invariants I_4, $I_6 - I_{32}$, when added to starting set (I_1, I_2, I_3 and I_5). Bad performance of individual invariants suggests strong correlation with the I_1, I_2, I_3, I_5 set. Middle bottom: the mean difference graph showing relative performance gain/loss for individual invariants. Note the relative decrease in performance for I_4 which completes the dependent set. Bottom: histogram of invariants selected in the process.

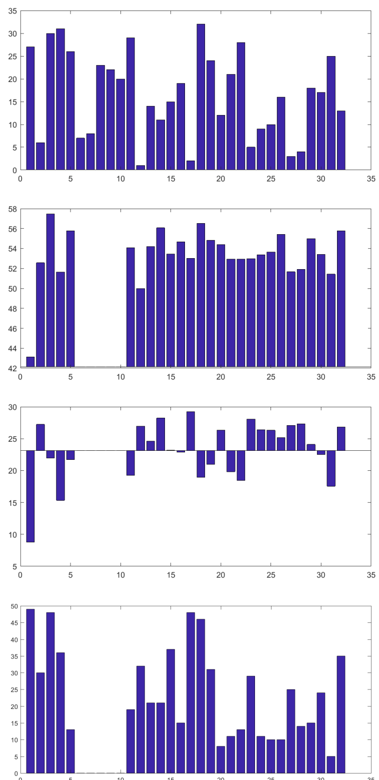

Fig. 6. Statistics on invariants discriminative powers when invoking d2 (5) dependency. Left: MNIST dataset, right: MEW 2014; top: the mean classification accuracies for individual invariants $I_1 - I_{32}$; middle top: the mean classification accuracies for invariants $I_1 - I_5$, $I_{11} - I_{32}$ when added to starting set $(I_6 - I_{10})$. Bad performance of individual invariants suggests strong correlation with the starting set. Middle bottom: the mean difference graph showing relative performance gain/loss for individual invariants. Note the decrease of I_1 performance due to the dependency to the starting set $I_6 - I_{10}$. Bottom: histogram of invariants selected in the process.

In the result, dependent feature sets were selected in 12/200 (6%) cases on MNIST database and 51/200 (25.5%) dependent feature set selected on MEW 2014. See Figure 5 bottom for histogram of both datasets. This corresponds to I_4 having greater relative discriminability in MEW 2014, or more correlated in MNIST dataset. The middle bottom images of Figure 5 show the relative changes in individual invariants performance and indicate the strong correlation of invariants in relation to the starting set.

Dependency d2 The setup for testing the strength of d2 in (5) was the same, only we initiated Feature Selection with $\{I_5, I_6, I_7, I_8, I_9\}$.

Invoking dependency d2 resulted in 60/200 (30%) cases of selecting I_1 for MNIST dataset and 52/200 (26%) for MEW 2014, showing significant discriminative power of this invariant. The clear I_1 significance declination is depicted in middle bottom images in Figure 6.

5 Conclusion

In our work, we have shown the importance of studying the theoretical dependencies between affine moment invariants and their direct impact on the classification performance. Our experiments show that the chance of generating set of affine moment invariants with polynomial dependency by the means of feature selection processes is less or equal to 2% in all cases. Some of these cases were positively identified as a situation where, during feature selection process, there were more invariants with the same classification performance, so the one with the smaller index was chosen, hence producing the dependent set. Those cases can be considered random noise.

In the second part of our experiments, we focused on the strength of dependencies d1 and d2 by removing most significant invariant from the dependent sets. We have confirmed that although it was possible to forcefully invoke independent set of affine moment invariants, this occurred in less than 30% of the cases, again showing power of the invariant dependencies in accordance to theory prediction.

This signifies that studying of the affine invariant dependencies have its purpose and is to be considered when using invariants as image features in pattern recognition tasks.

Such a study was never done before and up until now it was not clear, whether the theoretical properties of affine moment invariants have any noticeable relations to practical classification tasks in discrete computer domain, where continuous domain relations does not have to necessarily hold, due to the discretization process, numerical instabilities or even precision limitations of the computation.

In our ongoing work, we would like to broaden this study to include invariants of higher orders and confirm, that the same holds also for those more complex features. We would also like to be able to further study the numerical properties of discrete affine moment invariants in relation to calculation and usage in classification optimization processes.

References

1. Flusser, J., Suk, T.: Pattern recognition by affine moment invariants. Pattern Recognition **26**(1) (1993) 167–174
2. Suk, T., Flusser, J.: Affine moment invariants generated by graph method. Pattern Recognition **44**(9) (2011) 2047–2056
3. Flusser, J., Suk, T., Zitová, B.: 2D and 3D Image Analysis by Moments. John Wiley & Sons (2016)
4. Flusser, J., Zitová, B., Suk, T.: Moments and moment invariants in pattern recognition. John Wiley & Sons (2009)
5. Suk, T., Flusser, J.: Tables of affine moment invariants generated by the graph method. Technical report, Research Report 2156, Institute of Information Theory and Automation (2005)
6. Pudil, P., Novovičová, J., Kittler, J.: Floating search methods in feature selection. Pattern Recognition Letters **15**(11) (1994) 1119–1125
7. Vapnik, V.: The nature of statistical learning theory. Springer Science & Business Media (2013)
8. LeCun, Y., Cortes, C., Burges, C.J.: The MNIST database of handwritten digits (1998) `http://yann.lecun.com/exdb/mnist`, (accessed June 7, 2017).
9. Novotnỳ, P., Suk, T.: Leaf recognition of woody species in Central Europe. Biosystems Engineering **115**(4) (2013) 444–452
10. Suk, T., Novotný, P.: Middle European Woods (MEW 2012, 2014) (2014) `http://zoi.utia.cas.cz/node/662`, (accessed June 7, 2017).