Online reconstruction-free single-pixel image classification

Pedro Latorre-Carmona[a,*], V. Javier Traver[a], J. Salvador Sánchez[a], Enrique Tajahuerce[b]

[a]*Institute of New Imaging Technologies, Universitat Jaume I. Campus Riu Sec s/n, 12071. Castellón de la Plana, Spain*
[b]*GROC UJI, Institute of New Imaging Technologies, Universitat Jaume I. Campus Riu Sec s/n, 12071. Castellón de la Plana, Spain*

Abstract

In single-pixel imaging, a series of illumination patterns are projected onto an object and the reflected or transmitted light from the object is integrated by a photodetector (the single-pixel detector). Then, from the set of received photodetector signals, the image of the object can ultimately be reconstructed. However, this reconstruction is not only computationally expensive, but also unnecessary for purposes such as image classification tasks. This work proposes a reconstruction-free multi-class image classification framework that, unlike most of the existing approaches, exploits the sequential nature of the problem. Indeed, by accumulating evidence of the sequence of scalar values, a decision is made after each measurement on whether already classifying the object being imaged, or waiting for more measurements. This online decision relies on a mechanism to achieve a recognition-delay trade-off that induces behaviours within the conservative-to-aggressive spectrum, which suit distinct requirements in different applications. Additionally, the presentation order of the illumination patterns makes a difference in terms of the reconstruction quality (if required) and classification performance when a limited number of patterns is used. Nevertheless, in many cases simple data- and task-agnostic orders, such as random or frequency-based orders, are commonly used. To address this, a novel sparse-representation-based strategy is presented that sorts the patterns according to their general and discriminability utilities. Both, the online classification framework including the recognition-delay trade-off mechanism, and the data- and task-aware pattern ordering proposed, are experimentally assessed, with encouraging results, on the MNIST digits and CalTech 101 Silhouettes datasets.

Keywords: Single-pixel Imaging, Image classification, Recognition-delay trade-off, Sparse representations

1. Introduction

Single-pixel imaging (SPI) [1, 2] is a relatively recent image acquisition technique that offers several advantages in relation to conventional matrix-like camera devices. Probably, the most important ones relate to cost efficiency regarding the development of matrix sensors for some electromagnetic spectral ranges. However, these advantages come naturally with some drawbacks (e.g. acquisition time increases) and new challenges (e.g. generation of high-dimensional modalities such as multi-spectral imaging).

Although acquiring a (good enough quality) 2D conventional image from single-pixel measurements has been the mainstream concern and procedure [3], in many practical visual tasks the purpose is not obtaining the image itself but making some prediction, such as classifying the image into one among a given set of predefined classes. It is therefore desirable to perform these tasks directly from the photodiode measurements, with no image reconstruction at all. Despite its theoretical interest and practical relevance, this reconstruction-free problem has been comparatively much less addressed, and therefore motivates this work. Successful classification in the compressed domain (i.e. the set of photodetector readings) as opposed to the original (image) domain, has been reported [4], and theoretical bounds on the number of measurements to be made for reliable classification have been given for classification [5] and reconstruction [6].

Since measurements in single-pixel scenarios are performed sequentially, it is particularly interesting to explore mechanisms that might allow for online deciding after which measurement the classification of a given new target object can be performed reliably enough. Certainly, image reconstruction could be done incrementally and classification be performed in the image domain (or in the domain of features computed from images). Nevertheless, given that our goal is not image formation, it would be ideal to formulate a solution that bypasses image reconstruction, and performs classification in the measurement domain. Furthermore, not reconstructing the image is computationally advantageous not only because the reconstruction itself is not performed, but also because the classification can be performed in a (much) lower dimensional space.

1.1. Related work and contributions

From the design goals suggested above, our specific contributions are as follows:

- *A novel framework for online decision and a mechanism for a delay-performance trade-off.* The recognition strategy builds on multivariate Gaussians [7],

---

*Corresponding author
Email address:* `latorre@uji.es` (Pedro Latorre-Carmona)

and the decision itself partially relies on the sequential hypothesis ratio [8], but other formalisms could be used with minor changes on the proposed framework. The proposed online decision method considers a delay-recognition trade-off through time-pressure parameters. Although similar trade-off mechanisms have been considered elsewhere, they have been addressed for different problems and applications, such as action recognition [9], and with standard imaging devices rather than single-pixel devices.

- *A reconstruction-free decision framework.* Some approaches for single-pixel image classification require image reconstruction [1], but others skip this step and try to perform the classification directly from the measurements of the single-pixel photodetector [7, 5], or by convolving the photodetector signal with the so-called *smashed* filters, which are matched filters in the compressive-sensing scope [10, 1]. In essence, smashed filtering aims at classifying a signal by correlating it with the signal that characterises the class, directly in the low dimensional projection (compressive) space. The limitations of this linear filtering can be addressed with convolutional neural networks [11]. Recently, a naive Bayes classifier is applied on SPI using Fourier-based sinusoidal patterns and 1D Gaussian distributions [12] instead of the multidimensional Gaussians considered before [7]. In the context of both reconstruction and classification in the compressive domain, the use of side information is shown to improve performance [13].

- *Exploration of task-specific measurements.* We study how the order of the measurements (patterns) can be tailored for the specific classification tasks, based on available training data. To this end, we consider the single-pixel image reconstruction process as one of expressing the images as a *sparse* representation of the atoms in a dictionary. This relates to approaches which optimise either the presentation order of the illumination patterns [14] or the patterns themselves [15], or may adapt the projection space given past measurements [7]. It has been shown that data-agnostic random projects are outperformed by supervised data-dependent projections [16], which uses a concept resembling the dissimilarity space [17]. Recent deep-learning approaches [18] outperform smashed filters.

To summarise, and generally speaking, despite recent progress in compressed learning, many works either do image reconstruction, or just focus on the projected signal, but disregard the sequential nature of this signal, or do not consider alternative illumination patterns or their order. In contrast, our work addresses both the reconstruction-free and online classification, and explores alternative measurement orders as well.
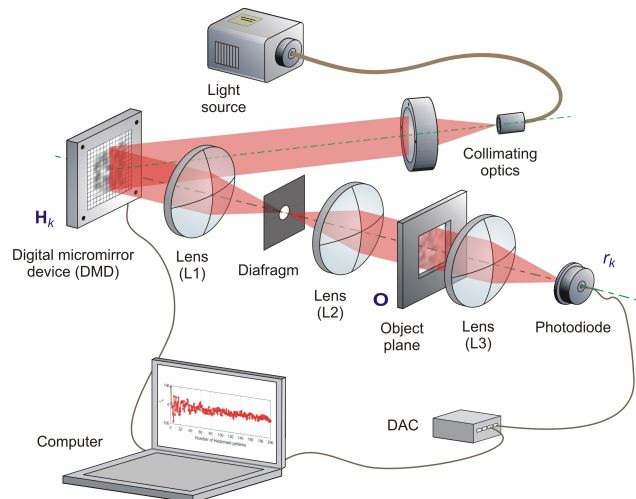

Figure 1: Schematic diagram of a single-pixel camera set-up. See text for an explanation. Although typically the reconstruction is required, we address the problem of recognizing the image class without its reconstruction and sequentially, i.e. by simply processing the photodiode readings for each projection pattern

The present paper is organised as follows. To make the presentation more self-contained, a succinct review of single-pixel imaging is first given (Sect. 2). Then, the proposed online classification techniques and strategies proposed for SPI are presented (Sect. 3) and carefully evaluated and compared (Sect. 4). Finally, conclusions (Sect. 5) are drawn.

2. Single-pixel imaging fundamentals

Single-pixel imaging techniques are based on illuminating the scene with a sequence of light structured patterns while the light reflected or transmitted by the objects is recorded by a single photodetector such as a photodiode or a photomultiplier tube [1]. In a typical configuration (Fig. 1), a digital micromirror device (DMD) is illuminated with the light produced by a white-light source, conveniently expanded, homogenised and filtered. Afterwards, the micro-structured light patterns codified on the DMD display are projected onto the object using a 4-f optical system formed by two lenses. A circular diaphragm at the focal plane filters unwanted diffracted orders. Then, the light transmitted by the object is collected and focused by a lens onto the photodetector. The correlation level between each light pattern and the object directly determines the photodiode electrical signal.

In particular, let us consider a structured light pattern formed by 2D functions $\mathbf{H}_k$ pertaining to the orthonormal Walsh-Hadamard (WH) basis: $\mathbf{H}_k$ are binary matrices whose elements are equal to $+1$ or -1. By representing the elements of these WH matrices as $H_k(m,n), m \in \{1,\ldots,M\}, n \in \{1,\ldots,N\}, k \in \{1,\ldots,M \cdot N\}$, the mea-

surement given by the photodiode can be modelled by

$$r_k = \sum_{m=1}^{M} \sum_{n=1}^{N} H_k(m,n) \cdot O(m,n), \qquad (1)$$

where $O(m,n)$ refers to the content of the object ($\mathbf{O}$) at position (m,n), and r_k represents the intensity measured by the photodetector for the k-th WH pattern. Due to the properties of these matrices, the image corresponding to the object scene, up to the k-th WH pattern, can be reconstructed as follows:

$$\tilde{O}(m,n) = \sum_{k=1}^{K} r_k \cdot H_k(m,n), \qquad (2)$$

where $\tilde{O}(m,n)$ would be the reconstructed approximation of $O(m,n)$ when using $K < M \cdot N$ patterns.

Fig. 2 shows an example of the reconstruction of an image formed by an instance of the digit 7. The first two rows show the reconstructed images using varying number k of patterns, for two different presentation orders, *zig-zag* (the WH patterns are presented in order of increasing frequencies) and random order, respectively. The third row in Fig. 2 shows the value, r_k, that would be given by the photodetector, for each one of the k patterns. Notice that it is this sequence of values $\{r_k\}_k$ for a given object, not its reconstructed images, that is used to classify the object —and decide when to. In can also be observed that the presentation order of the illumination patterns has an impact on the quality of the reconstructed image for a given number of patterns K and, arguably, on the confidence of early classification based on r_k. For this reason, besides the conventional frequency-based (*zig-zag*) ordering, an alternative ordering is proposed (Sect. 3.2).

3. Methodology

From a theoretical point of view, the central idea that allows for detection or recognition problems to be solved directly in the low-dimensional projective space is the Johnson-Lindenstrauss (JL) lemma and its application in *smashed filtering* [5, 19]. Specifically, the JL lemma states that the correlation between any two signals is nearly preserved even when the data is compressed to a much lower-dimensional space.

Our approach for the reconstruction-free online classification relies on estimating the classification confidence from the past optical measurements. In essence, the more measurements are made, the higher the classification confidence. To that end, we follow [7], but instead of using as many patterns as necessary to classify a novel instance with a classification error below a given threshold, we propose a framework for online classification where decisions are made as measurements are being received sequentially following a recognition-delay trade-off policy. The estimation of the classification confidence assumes a multivariate

Gaussian distribution in the projected space of the original observations (measured patterns).

Taking into account (1), we can define $\mathbf{h}_k \in \mathbb{R}^{K \times 1}$ as the lexicographically ordered version of $\mathbf{H}_k$ and $\mathbf{H}^{(k)} = [\mathbf{h}_1, \ldots, \mathbf{h}_k]$ and $\boldsymbol{r}_k = [r_1, \ldots, r_k]$. Considering $\mathbf{o} \in \mathbb{R}^{K \times 1}$ as the lexicographically-ordered version of $\mathbf{O}$, the object probability density function is given by

$$P(\mathbf{O}|C_i) = \frac{\exp\left(-\frac{(\mathbf{o}-\boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1}(\mathbf{o}-\boldsymbol{\mu}_i)}{2}\right)}{(2\pi)^{\frac{N}{2}} |\boldsymbol{\Sigma}_i|^{\frac{1}{2}}}, \qquad (3)$$

with C_i being the hypothesised class i. The mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$, for each class $i \in \{1, \ldots, Q\}$, are computed from the corresponding training subset $\{\mathcal{S}_i\}_1^Q$, each of size S_i,

$$\boldsymbol{\mu}_i = \frac{1}{S_i} \sum_{j=1}^{S_i} \mathbf{o}_{ij}; \quad \boldsymbol{\Sigma}_i = \frac{1}{S_i} \sum_{j=1}^{S_i} \left[(\mathbf{o}_{ij} - \boldsymbol{\mu}_i)(\mathbf{o}_{ij} - \boldsymbol{\mu}_i)^T\right], \qquad (4)$$

with $\mathbf{o}_{ij}$ being the object in j-th training instance in the set $\mathcal{S}_i$, $j \in \{1, \ldots, S_i\}$.

Taking (4) and (3) into account, the a-posteriori probability distribution is given by

$$P(\boldsymbol{r}_k|C_i) = \frac{\exp\left(-\frac{(\boldsymbol{r}_k-\boldsymbol{\mu}_i^k)^T [\boldsymbol{\Sigma}_i^k]^{-1}(\boldsymbol{r}_k-\boldsymbol{\mu}_i^k)}{2}\right)}{(2\pi)^{\frac{kL}{2}} |\boldsymbol{\Sigma}_i^k|^{\frac{1}{2}}}, \qquad (5)$$

where $\boldsymbol{\Sigma}_i^k = (\mathbf{H}^{(k)})^T \boldsymbol{\Sigma}_i \mathbf{H}^{(k)} + 2\sigma^2[\mathbf{I}]$ and $\boldsymbol{\mu}_i^k = (\mathbf{H}^{(k)})^T \boldsymbol{\mu}_i$, with $\mathbf{I}$ being the identity matrix.

Rather than simply classifying when a likelihood of probabilities is higher than a given threshold, as in [7], a richer decision function is used in our work since the proposed approach includes a recognition-delay trade-off (Sect. 3.1).

It is illustrative to observe the evolution of (5) with increasing number of patterns received k. When evaluated on the instances in a particular class in the test set for two image datasets (Fig. 3), the general trend is that this probability grows rapidly with k for the true class, while remains constant and lower for the rest of the classes. Not surprisingly, this result depends on the class and the dataset, and some class overlaps may also happen. The purpose of the proposed recognition-delay trade-off is precisely to capture, formalise, and exploit the observation that just after very few measurements, the true class may usually be predicted with reasonable confidence.

3.1. Online decision with delay-recognition trade-off

Given the above observations, the problem can be stated as follows: after having performed the k-th measurement for a particular (test) instance, decide whether we already have enough confidence to classify this instance, or it is preferable to wait and perform some more measurements before. Furthermore, depending on the target application, one can take higher or lower risk in an early classification.
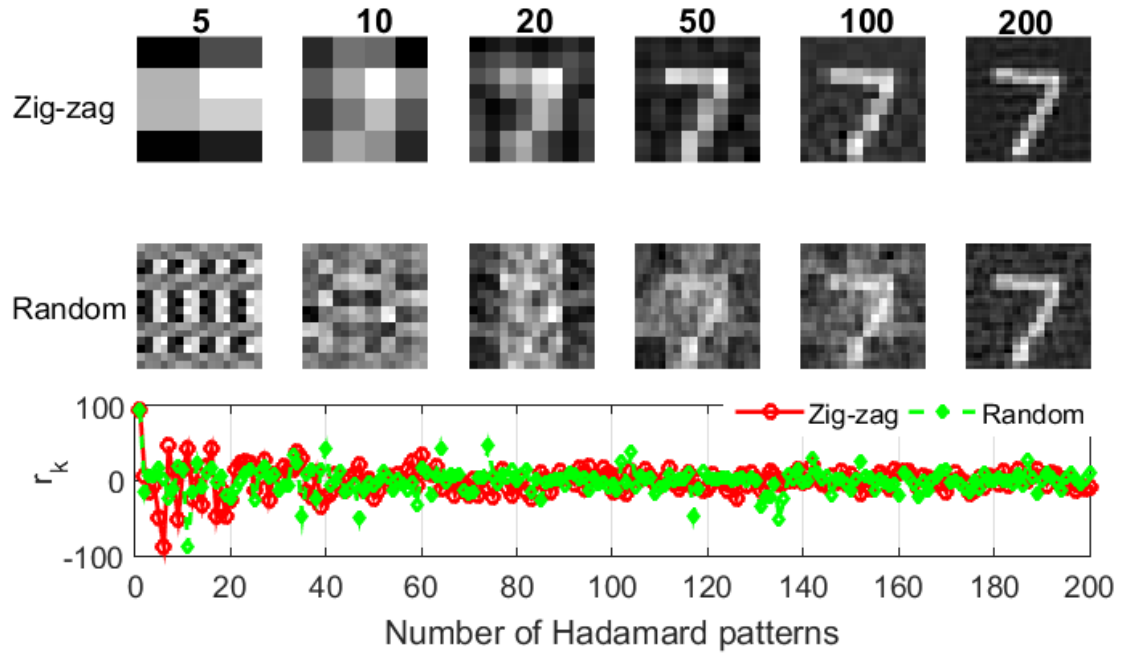
Figure 2: Reconstruction of the image of a digit 7 from the MNIST dataset and the corresponding photodetector signal r_k for $k \in \{5, 10, 20, 50, 100, 200\}$ Walsh-Hadamard (WH) patterns. Two different presentation orders for the patterns have been used: *zig-zag* and random. Notice that reconstruction is shown only for comparing the underlying pattern order, since the proposed online strategies do not perform reconstruction
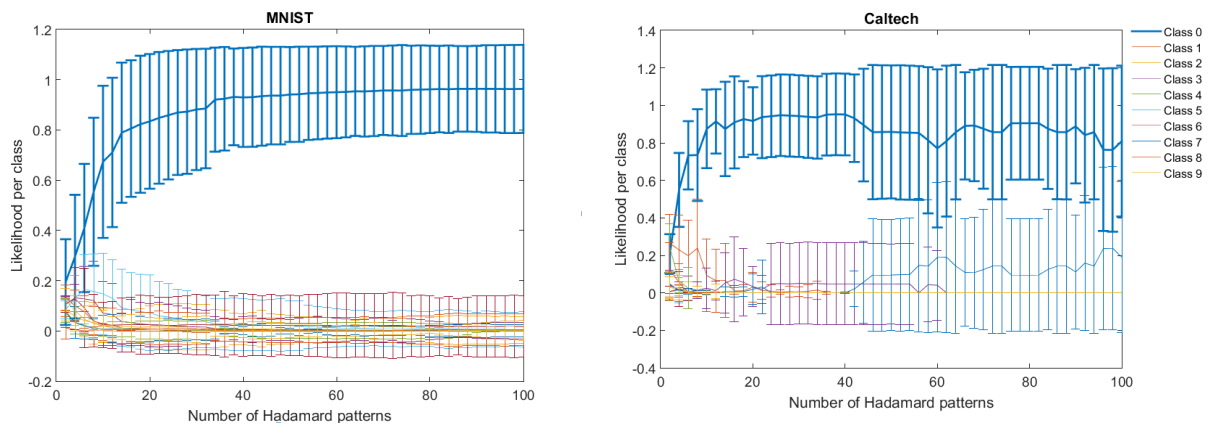

Figure 3: Mean and standard deviation of the a posteriori class-conditioned probabilities, Eq. (5), as a function of the received patterns for one class (Class 0) of the MNIST and CALTECH datasets (Fig. 5). The frequency-based order is used in this case

To formalise these notions, we define an evidence function f that consists of two terms: r, which models the classification (recognition) confidence, and d, which represents the relative time passed (delay),

$$f(t) = (1 - \lambda) \cdot r(t) + \lambda \cdot d(t), \qquad (6)$$

with $\lambda \in [0, 1]$ being a time-penalization factor to emphasise the relative importance given to early classification (higher λ) or to more confident classification (lower λ).

The recognition r and delay d functions are defined as follows:

$$\begin{aligned} r(t) &= 1 - \exp\left(-\kappa \cdot \frac{P_1}{P_2}\right), \\ d(t) &= 1 - \xi \cdot \exp\left(-\sigma \cdot \frac{T-t}{T}\right). \end{aligned} \qquad (7)$$

where P_1 and P_2 represent, respectively, the largest and second-largest probabilities in (5) among all possible classes, considered at time t; t is a measure of the relative time (here, $t = k$, i.e. the number of measurements performed so far); and $T = K$, the total number of patterns.

While λ plays a bigger role for the recognition-delay trade-off, κ also has an influence: the lower the value for κ the lower $r(t)$ and, in turn, the lower $f(t)$. This implies that, even for the same λ, classification may also be delayed with lower κ. Thus, λ and κ govern the problem-dependent decision strategy so that more aggressive or conservative policies can be defined. Together, both parameters model what we refer to as the *time pressure*. As for the other free parameters in (7), they were empirically set to $\xi = 0.3$ and $\sigma = 9$.

The decision made after each pattern k is measured consists of either classifying the image as class

$$\hat{c} = \arg\max_{C_i \in \{1,\dots,Q\}} P(\boldsymbol{r}_k | C_i),$$

with probability f, or postponing the classification and therefore performing yet another new measurement with probability $1 - f$. However, to make this decision more robust, the actual classification for a particular instance is made only after an accumulated minimum number n of times the classification has been chosen for that instance. In this work, $n = 3$ was used.

3.2. On the pattern order: can we do better?

One very simple order of presenting the WH patterns is the so-called random order (RO), where the patterns are selected uniformly at random. Another approach consists of scanning the $2D$ frequency space of the WH space in a so-called *zig-zag* order (from the lowest to the highest frequencies). We will also refer to this order as the frequency-based order (FO) from now on. Although the frequency-based order is in many cases considered a default reasonable choice, at least for image reconstruction purposes, alternative orders might be more adequate for other problems such as classification. In particular, here we explore an strategy based on finding sparse representations of the

images by defining a dictionary of all the possible illumination patterns. Given the dictionary $\mathbf{D}$ of K atoms (the WH patterns), the coefficients $\boldsymbol{\alpha}$ are chosen to sparsely represent each training image $\mathbf{O}_i$ in terms of the dictionary atoms so that $\mathbf{o}_i$, the lexicographically-ordered version of $\mathbf{O}_i$, would be $\mathbf{o}_i \approx \mathbf{D}\boldsymbol{\alpha}_i$, with $\mathbf{o}_i \in \mathbb{R}^{K \times 1}$, $\mathbf{D} \in \mathbb{R}^{K \times K}$ and $\boldsymbol{\alpha}_i \in \mathbb{R}^{K \times 1}$, $i \in \{1, \dots, S\}$.

Thus, the sparse coefficients for the S images $\mathbf{X} = [\mathbf{o}_1, \dots, \mathbf{o}_S] \in \mathbb{R}^{K \times S}$ in the training set are found by minimizing the reconstruction error subject to a maximum number $\boldsymbol{\tau}$ of non-zero coefficients, i.e.

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{D} \cdot \mathbf{A}\|_2^2 \quad \text{s.t.} \quad \|\boldsymbol{\alpha}_i\|_0 \leq \boldsymbol{\tau}, \qquad (8)$$

with $\mathbf{A} = [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_S] \in \mathbb{R}^{K \times S}$. This optimization problem is solved by Orthogonal Matching Pursuit (OMP) [20]. In particular, the OMP-Box v10 implementation [21] based on [22] was used here. A value of $\boldsymbol{\tau} = 100$ was considered in all cases.

Now, we decide on the *utility* of the WH patterns based on the corresponding coefficients. Intuitively, the larger the absolute values of the sparse coefficients associated to a given pattern, the more useful this pattern generally is because the reconstruction quality of the images rely more on this pattern than on others with lower sparse coefficients. On the other hand, for classification purposes, this general utility may not be very relevant and we might prefer those patterns that induce higher class separability. To account for both objectives, the utility U_j of pattern j is assessed by the composition of a general utility G and a discriminability D terms,

$$U_j = G_j + w_d D_j, \qquad (9)$$

with $w_d \geq 0$ providing a relative emphasis of discriminability over the general utility. We analysed the minimum and maximum values that both terms (G_j and D_j) could reach, and found that the maximum values of D_j could be an order of magnitude bigger than that of G_j. We therefore tested the performance of U_j for different values of $w_d \in [0, 1]$ and found that the best results were obtained with $w_d = 1$, which was thus used for the reported experiments.

Although there are several possible definitions for G, it was found the following one to perform reasonably well:

$$G_j = \text{median}_i\{|\alpha_{ij}|\}, \quad i \in \{1, \dots, S\}, j \in \{1, \dots, Q\}, \qquad (10)$$

with α_{ij} being the sparse coefficient of the training instance i corresponding to pattern j, and $S = \sum_{i=1}^{Q} S_i$ is the total number of training instances used in OMP. The median was observed to perform better than the mean.

As for D, it can be modelled as the sum of pair-wise

class discriminability,

$$D_j = \sum_{\substack{\{a,b\} \in \{C_i\}_{i=1}^Q, \\ a \neq b}} D_j(a,b), j \in \{1,\ldots,Q\}. \quad (11)$$

By denoting G_j^c as G_j computed marginally for a particular class c (i.e. using instances only in the set $\mathcal{S}_c$), then the discriminability between two particular classes a and b, $D_j(a,b)$, can in turn be expressed in terms of G_j^c as

$$D_j(a,b) = |G_j^a - G_j^b|, j \in \{1,\ldots,Q\}. \quad (12)$$

Following this approach, the presentation order of the patterns is given by decreasing values of their utilities U_j, i.e. $\{j_1, j_2, \ldots j_K\}$, $U_{j_r} \leq U_{j_s}, r < s, r, s \in \{1,\ldots,Q\}$.

We refer to the order induced by this utility as the sparse-based utility order (SUO), and will be compared against the conventional frequency-based order (FO).

As an illustrative example of the potential usefulness of SUO over alternative orders, the reconstruction for an image of the digit 2 (Fig. 4) from the MNIST dataset with random, FO, and SUO orders, with increasing number k of Hadamard patterns, suggests that the RO is the worst among the three, and the reconstructed image has higher quality *earlier* with SUO than with FO: with $k = 20$ patterns the digits is already quite recognisable under SUO, but not yet with FO. Actually, with 10 or even 5 patterns, the reconstructed images capture better the main traits of the digit with SUO than with FO. This suggests that the corresponding photodetector measurements might also help discriminate images of different classes quicker.

4. Experiments

As an experimental testbed before testing the proposed framework on an actual hardware setup in the future, the single-pixel imaging reconstruction is simulated. This is done by considering the object **O** as a digital image and applying (2) for the desired number K of patterns.

After describing the datasets (Sect. 4.1) and the baselines (Sect. 4.2), the proposed approaches regarding recognition-delay performance (Sect. 4.3), and the pattern presentation order (Sect. 4.4) are evaluated.

4.1. Datasets

Two datasets were chosen for experimentation: MNIST and CalTech, described below. In both cases, the images sizes considered and the complexity of their contents are similar to those typically considered in experiments in optically implemented set-ups of single-pixel imaging.

The MNIST dataset of handwritten digits [23, 24] (MNIST from now on) has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger set available from the National Institute of

Figure 5: One instance per class from MNIST (above) and CalTech (below)

Standards and Technology (NIST). The digits have a fixed size of (28×28) pixels and its corresponding pixel values range in $[0,1]$. This dataset contains images of the digits from 0 to 9 in different orientations and appearances. These images were re-scaled to 32×32 pixels and therefore the maximum number of Hadamard patterns to use is $K = 32^2 = 1024$.

The CalTech 101 28×28 *Silhouettes* [25, 26] (CalTech from now on) contains a series of binary images with the silhouettes of 101 different classes. We randomly selected 10 of these classes, without any a priori knowledge about their shapes or characteristics, and resized their images to 32×32 pixels each. The number of classes for CalTech was essentially decided to match that of MNIST. The resulting dataset had 885 images for training and 1,249 for testing. Looking at examples of images of the chosen datasets (Fig. 5), it can be argued that the considered classes in CalTech look more alike, and possibly harder to distinguish for a human observer than the different digits in MNIST. For instance, using the numbers of the digits to refer also to the classes in CalTech, silhouettes of classes 0 and 1 on the one hand, and those of classes 5 and 6 on the other hand, resemble significantly one another. Not surprisingly, this results in higher class overlap in CalTech than in MNIST as illustrated in the class likelihoods (Fig. 3). In any case, MNIST images are more stroke-like whereas CalTech images are more blob-like. This different nature of these datasets is interesting in order to explore our strategies in single-pixel scenarios under different conditions.

4.2. Baselines and sanity checks

Since the proposed methodology includes reconstruction-free classification and alternative patterns and/or ordering, the natural baselines to compare the proposed approaches with are reconstruction-based classification and the conventional orders of WH patterns (namely, random and frequency-based).

For the reconstruction-based classification, a Support Vector Machine (SVM) with a Radial-Basis Function (RBF) kernel was used, and the raw pixel values were chosen as features, previously normalised to zero-mean, unit-variance. We evaluate SVM on the test set by reconstructing the image after each of the first k patterns. Regarding the training data, one choice is to use the *full-quality* images (original, fully-reconstructed images), and we will refer simply as SVM to the SVM classifier trained like this. However, since partially reconstructed images look
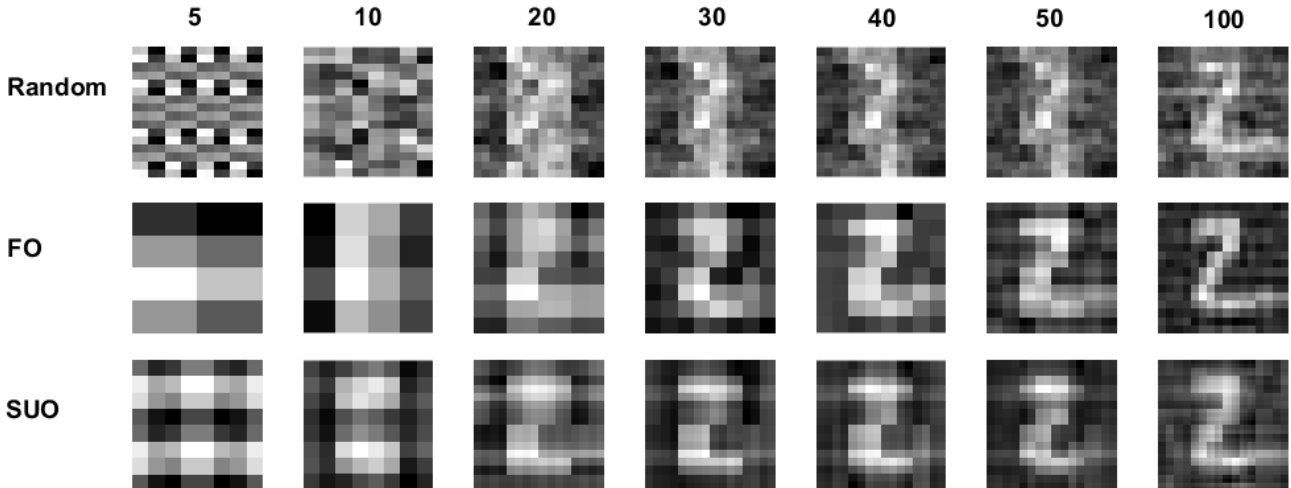
Figure 4: Reconstruction of an image of the digit number 2 of the MNIST dataset with the random, frequency-based (FO) and sparse-based utility (SUO) orders, for $k \in \{5, 10, 20, 30, 40, 50, 100\}$ number of Hadamard patterns.

like (very) differently to the fully reconstructed images, we wondered how SVMs specifically trained with data of a given reconstruction quality (i.e. after a given number of measures k), SVM_k, would perform when tested with the corresponding data. Although this idea was proposed in the framework of *online* stream data classification [27], it is the first time it is applied in single-pixel imaging, to the best of our knowledge.

We therefore trained separate $\mathrm{SVM}_{k_{tr}}$ for $k_{tr} \in \{10, 20, 50, 100, 500, 1,000\}$ patterns. A test instance reconstructed with the first k patterns is classified with $\mathrm{SVM}_{k_{tr}}$ with the smallest k_{tr} such as $k \leq k_{tr}$. Although this multi-SVM choice is computationally heavier due to multiple training procedures required, and its practical appeal is questionable, it is theoretically interesting.

SVM and SVM_k were trained using a 3-fold cross-validation strategy, considering all the elements in the corresponding training datasets. The selection of the best parameters for the SVM and SVM_k models was found applying a grid sampling strategy in the space of the classifier hyperparameters parameters (the common regularisation parameter and the scale parameter of the Gaussian of the RBF kernel).

Results (Fig. 6) indicate that, as expected, the more measures are performed, the higher the recognition performance, which agrees with the correspondingly increasing reconstruction quality (Fig. 2). Second, frequency-based order leads to faster convergence than random order, both in SVM and multi-SVM approaches. Finally, and very interestingly, the multi-SVM approach significantly outperforms the single SVM one. This means that training and testing on similar-quality images is more discriminative than training on full-quality images and testing on lower-quality (partially reconstructed) images.

4.3. Effect of the delay-recognition trade-off

To illustrate the effectiveness of the proposed mechanism for modulating the delay-recognition trade-off, for several time pressures (λ, κ), the number of both, correct classifications and misclassifications, were computed as the number of photodetector measurements were obtained. The results (Fig. 7) clearly illustrate that when the time pressure is small ($\lambda = 0, \kappa = 0.01$) the system behaves *conservatively* and the number of classifications made increases more slowly with time, which also leads to a smaller number of wrong classifications (i.e. high recognition with significant delay). Analogously, as the pressure is increased, the system behaves more *aggressively*: more classifications are made earlier, which increases both the number of correct and wrong classifications. It can be seen that unless the time pressure is extreme, the number of misclassifications is kept moderately low. Which time pressure is adequate for a given problem is largely an application-dependent decision, and might be chosen by cross-validation in a validation dataset. It can also be observed (Fig. 7) that, as explained above (Sect. 3.1), even for the same λ, classification may also be delayed with lower κ.

To give a quantitative account of these curves (Fig. 7), the accuracy after $k = 130$ WH patterns for this same set of time pressures (Table 1) clearly illustrates the impact on the final recognition rate of the time pressures. At least for the values considered, the effect of increasing λ is higher than the effect of increasing κ. The effect of a larger variety of time pressures will be considered below when evaluating the pattern orders (Sect. 4.4).

4.4. On the presentation order for the patterns

To get a better understanding of the performance of the presentation orders (FO vs SUO), the recognition rates (R) and average delays (D) are studied for varying time pressures (λ, κ). The resulting R-D plots (Fig. 8) reveal that,
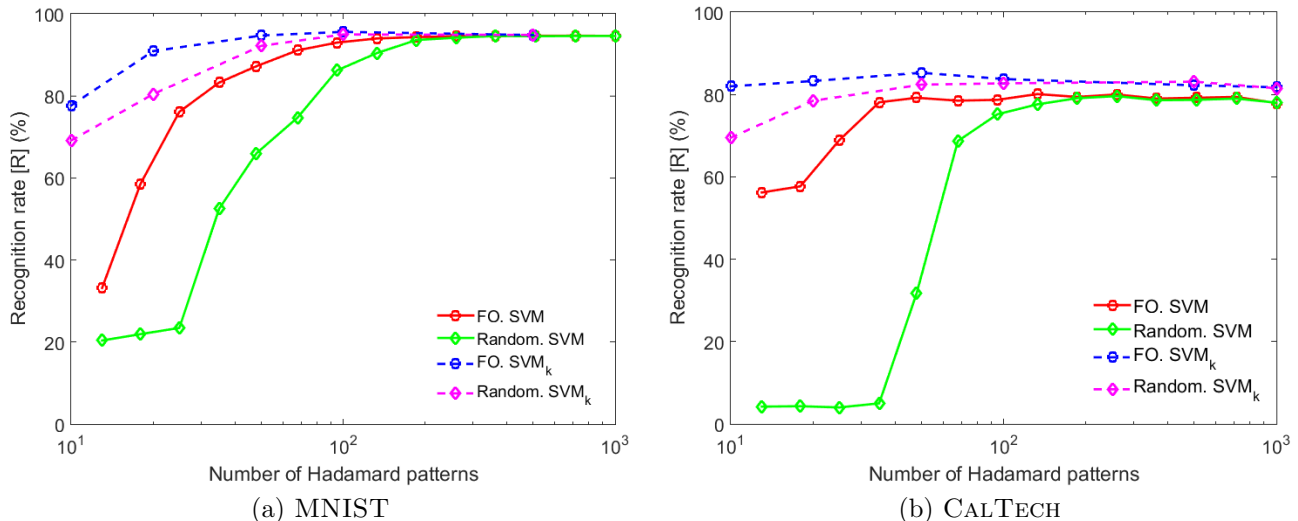
Figure 6: Classification accuracy for the random and frequency-based (FO) orders for the (a) MNIST and (b) CALTECH datasets, with SVM trained either on the original images (SVM) or on the reconstructed images for a specific number of patterns, k (SVM$_k$). Notice the delay (horizontal) axis is in logarithmic scale to better appreciate the different behaviours for the very first patterns. To facilitate the analysis of the plot, SVM (SVM$_k$) is represented with solid (dashed) lines, and FO (random) order with circles (diamonds)

Table 1: Accuracy (%) for different time pressures

(a) MNIST			(b) CALTECH		
	κ			κ	
λ	0.01	0.1	λ	0.01	0.1
0.0	92.5	82.6	0.0	79.1	78.7
0.5	66.3	63.7	0.5	73.3	72.4

as expected, the recognition rate is poor when early decisions are demanded through strong time pressures (with higher λ or κ or both). For softer time pressures, the recognition rate increases at the expense of longer delays. Reasonably, the variance in the delay (depicted here with the dot size) also increases with softer time pressures because different instances are classified at more different time steps, depending on the confidence in their classification.

When comparing FO and SUO for the same time pressure, different behaviours emerge in different datasets: in MNIST, SUO performs more aggressively than FO, i.e. SUO induces lower delay but also lower recognition; in CALTECH, lower delays and often higher recognition rates are obtained with SUO than with FO. The better performance of SUO in CALTECH than in MNIST might be partially attributable to the different nature of the images (blobs vs strokes), as observed before (Fig. 5). This indicates that although SUO does not systematically outperform FO, it is an option to consider in some cases (e.g. some datasets), and that there might be better choices than the commonly known and widely used frequency-based order.

By looking at these R-D plots, it can also be observed that the higher-performance region (upper-left corner, corresponding to lower delays and higher recognition rates),

can be reached by either FO and SUO for different time pressures. This means that the performance space can be better sampled not only with different time pressures, but also with different pattern order.

An interesting ramification of these observations is that, since FO and SUO induce distinct, complementary behaviours, a better strategy might be designed by combining both and getting the best of each.

4.4.1. Performance under noise

We also analysed how the proposed online classification strategy deals with noisy test images. When Gaussian noise is added, the pixel values are renormalised to their original [0,1] range. In the case of the CALTECH dataset, the images after noise injection are not binary anymore. Results for different levels of Gaussian noise (Fig. 9) show that, logically, the higher the noise, the lower the recognition rate and the higher the delay. It is particularly interesting the situation for the lower time pressure ($\lambda = 0, \kappa = 0.001$) and MNIST for both FO and SUO: R is minimally affected because the system implicitly discovers the higher difficulty in the images (higher class overlap) and postpones the classification decision more and more with increasing noise levels. This nicely illustrates the adaptive behaviour emerging from the proposed strategies. As with the noise-free case, for many cases of time pressure and noise level considered, SUO outperforms FO in terms of R in the CALTECH dataset. Interestingly, in the case of MNIST, although FO performs better than SUO when compared under the same noise level, the noise tends to affect R less when using the SUO than with FO, which is particularly noticeable for the higher time pressures.

4.4.2. Comparison in reconstruction-based classification

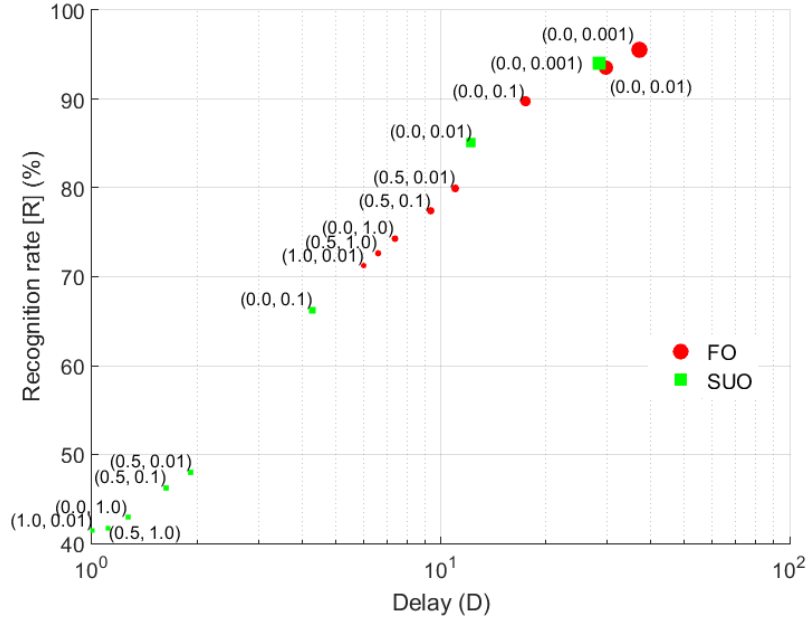Since SUO is not specific to the online scenario, it was
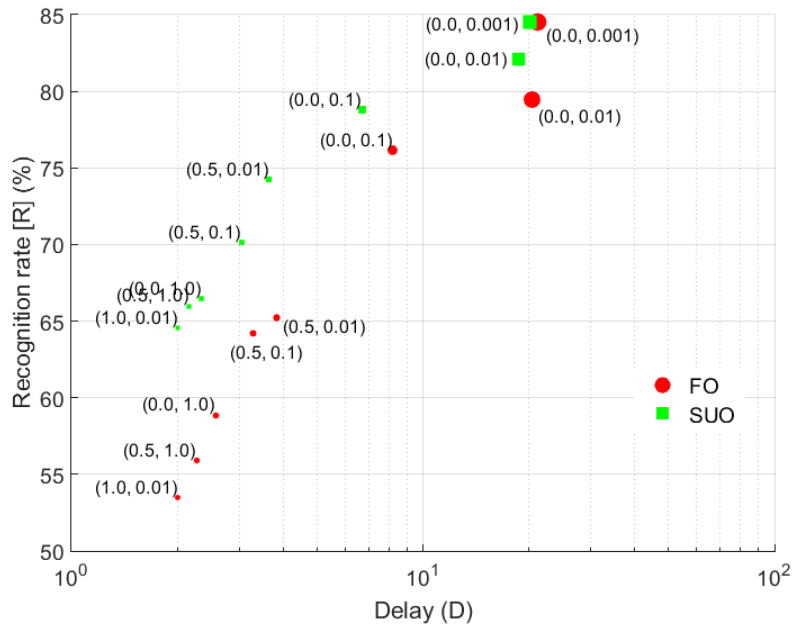
8
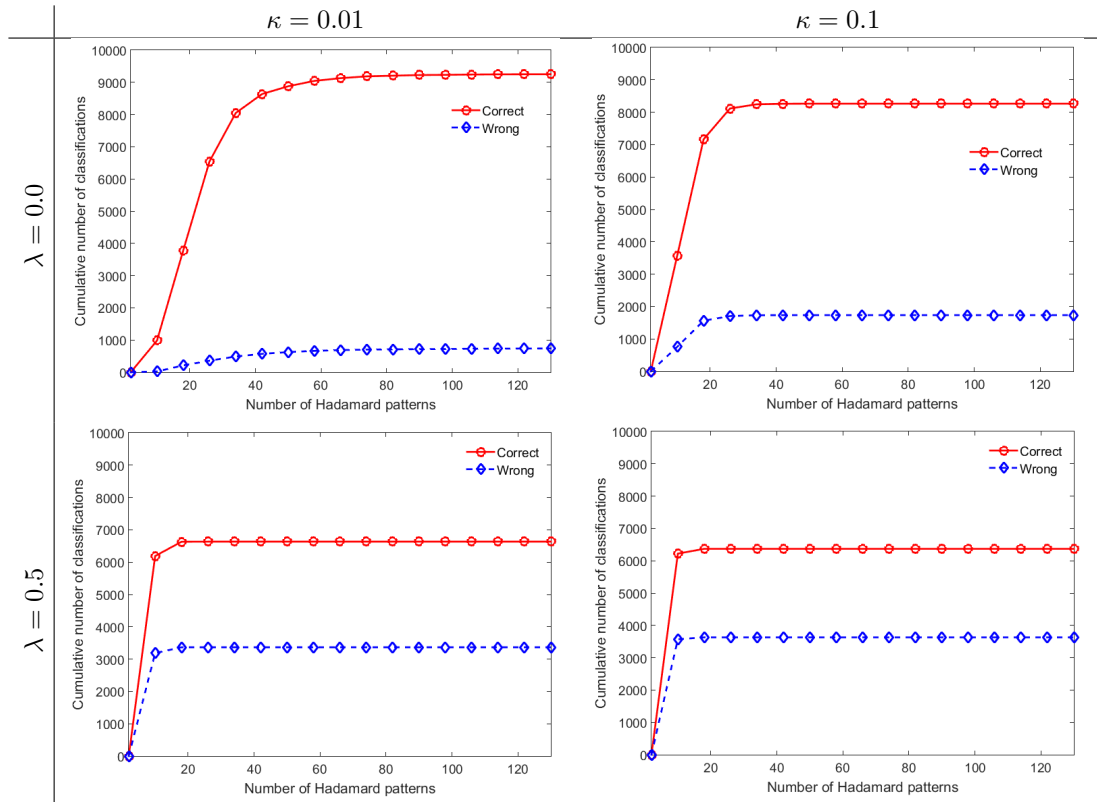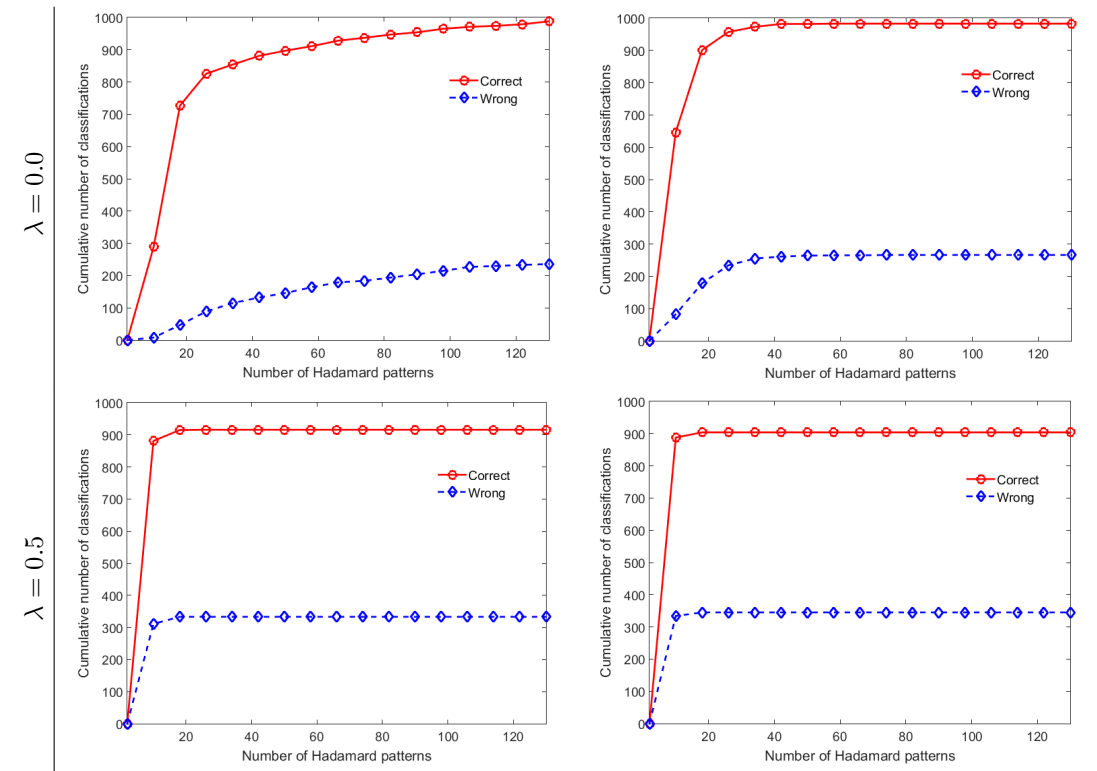
Figure 7: Evolution of the number of correctly and wrongly classified instances with increasing delay for (a) MNIST and (b) CALTECH, for different time pressures (λ, κ). For convenience, only one every 10 points are drawn

(a) MNIST

(b) CALTECH

Figure 8: Recognition rate R and delay D (in log scale) for FO and SUO by varying the time pressures (λ, κ) for (a) MNIST and (b) CALTECH. The dot sizes are proportional to the delay variance for each (λ, κ) pair. (Figure better seen in color)
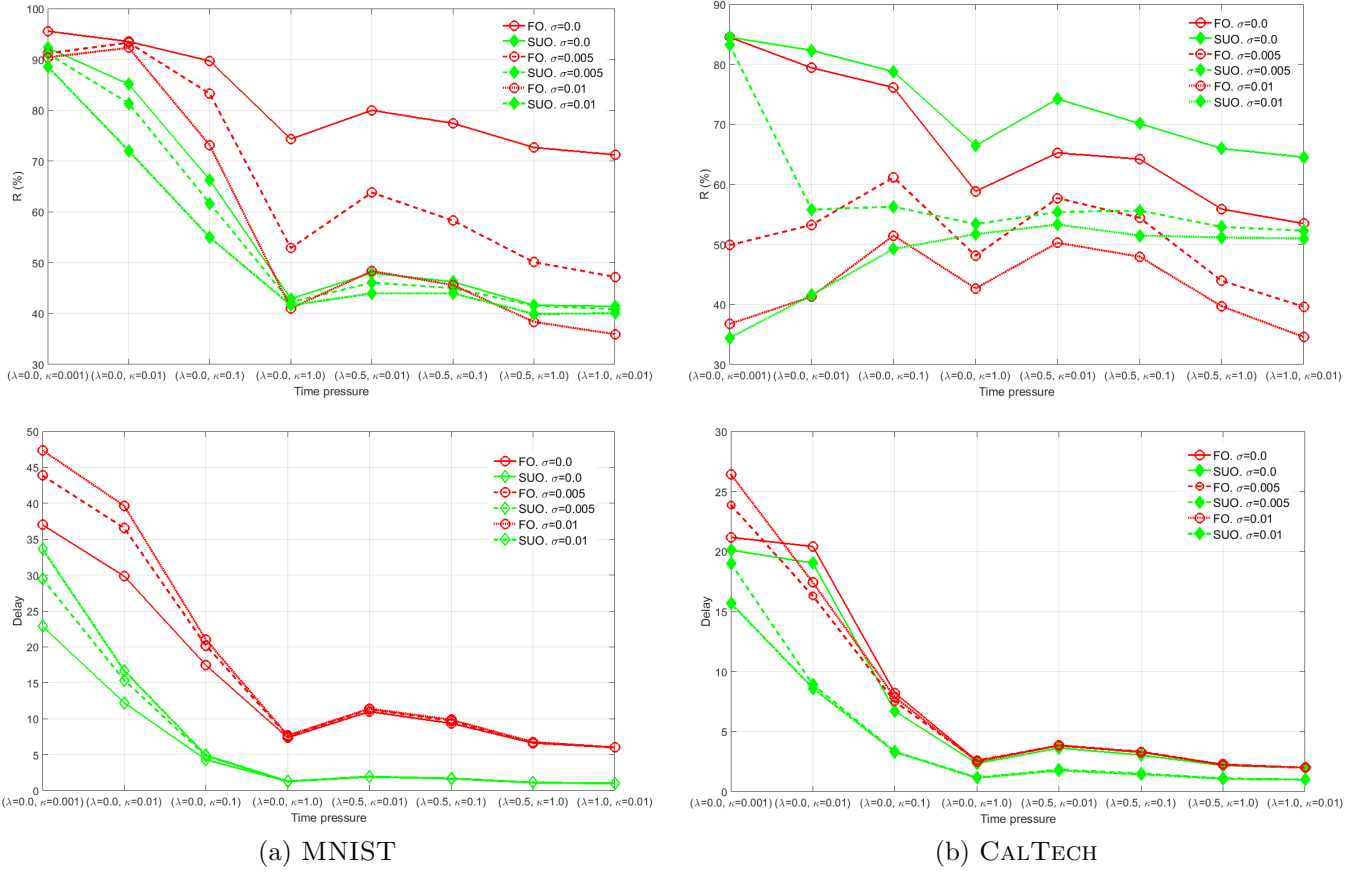
Figure 9: Recognition rate R (above) and delay D (below) with clean and noisy test images for different time pressures (λ, κ) for (a) MNIST (left) and (b) CALTECH (right). For the noisy versions, Gaussian noise was added with different standard deviation values $\sigma \in \{0.005, 0.01\}$; the clean case is denoted with $\sigma = 0$. For convenience, FO (SUO) is represented with circles (diamonds), whereas solid, dashed, and dotted lines represent increasing noise levels. Please, note that the jump in R and D from $\lambda = 0.0$ to $\lambda = 0.5$ can be better understood if the lines are considered unrelated (disconnected) at that point
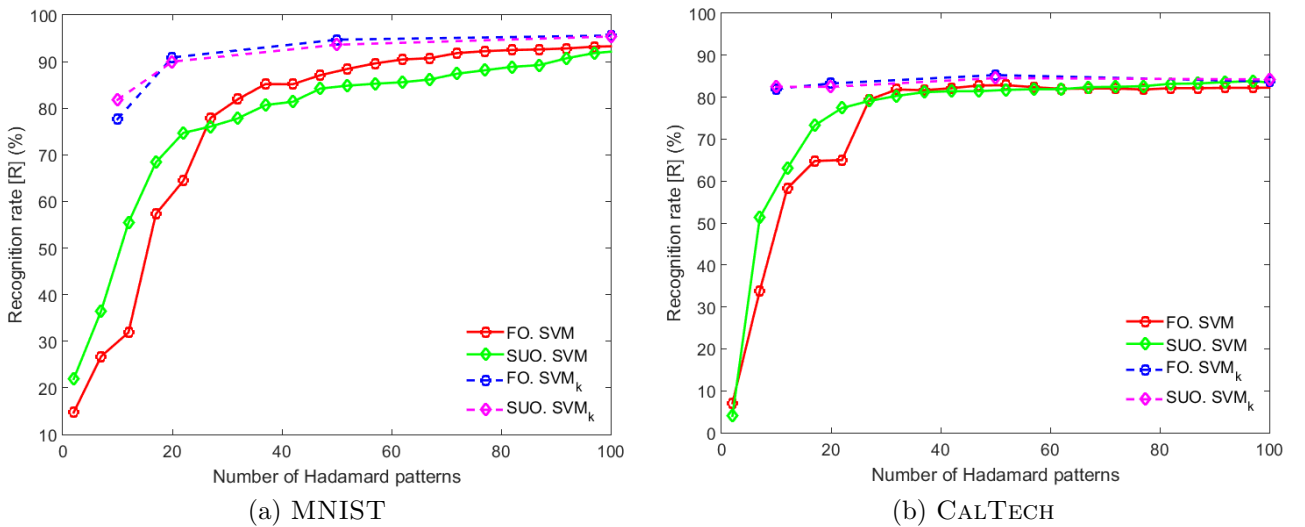

Figure 10: SVM accuracy on the MNIST and CALTECH datasets for the FO and SUO orders of the WH patterns, for the reconstruction-based case

11

also evaluated under the reconstruction-based classification, where it can be observed (Fig. 10) that SUO works better than FO when a small number of Hadamard patterns are to be chosen, both in MNIST and CalTech. When this number increases, the FO strategy outperforms SUO. This again calls for a proper synergetic usage of these different orders: as a very simple possibility, one may start with SUO and switch to FO after a suitable number of patterns.

5. Conclusions

Strategies for the online classification of objects imaged with the single-pixel formalism have been proposed which leverage the sequential nature of the problem. Experiments reveal that reconstruction-free classification, based only on the sequence of photodetector measurements, is possible, and that the proposed recognition-delay trade-off is effective. Additionally, the potential of a novel presentation order of the illumination patterns based on utility functions defined on their sparse representations, has been experimentally tested, and benefits observed in both, the reconstruction-based and the reconstruction-free cases.

Acknowledgments

References

[1] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, B. R. G., Single-pixel imaging via compressive sampling, IEEE Signal Processing Magazine 25 (2) (2008) 83–91.

[2] M. P. Edgar, G. M. Gibson, M. J. Padgett, Principles and prospects for single-pixel imaging, Nature photonics 13 (2019) 13–20.

[3] Y. Jauregui-Sánchez, P. Clemente, P. Latorre-Carmona, J. Lancis, E. Tajahuerce, Single-pixel imaging using photodiodes (Nov. 2018).
URL https://www.intechopen.com/online-first/single-pixel-imaging-using-photodiodes

[4] R. Calderbank, S. Jafarpur, R. Schapire, Compressed learning: universal sparse dimensionality reduction and learning in the measurement domain, Tech. rep. (2009).

[5] H. Reboredo, F. Renna, R. Calderbank, R. D. Rodrigues, Bounds on the number of measurements for reliable compressive classification, IEEE Transactions on Signal Processing 64 (22) (2016) 5778–5793.

[6] F. Renna, R. Calderbank, L. Carin, M. R. D. Rodrigues, Reconstruction of signal drawn from a gaussian mixture via noisy compressive measurements, IEEE Transactions on Signal Processing 62 (9) (2014) 2265–2277.

[7] P. K. Baheti, J. Ke, M. A. Neifeld, Adaptive feature-specific imaging for recognition of non-gaussian classes, Applied Optics 48 (28) (2009) 5225–5239.

[8] A. Wald, Sequential tests of statistical hypotheses, Ann. Math. Statist. 16 (2) (1945) 117–186.

[9] C. Ellis, S. Z. Masood, M. F. Tappen, J. F. LaViola Jr, R. Sukthankar, Exploring the trade-off between accuracy and observational latency in action recognition, International Journal of Computer Vision 101 (2013) 420–436.

[10] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, R. G. Baraniuk, The smashed filter for compressive classification and target recognition, in: Computational Imaging, 2007, p. 64980H.

[11] S. Lohit, K. Kulkarni, P. K. Turaga, Direct inference on compressive measurements using convolutional neural networks, in: IEEE International Conference on Image Processing, 2016, pp. 1913–1917.

[12] S. Jiao, Fast object classification in single-pixel imaging (2018). arXiv:1805.07582.
URL http://arxiv.org/abs/1805.07582

[13] L. Wang, F. Renna, X. Yuan, M. R. D. Rodrigues, A. R. Calderbank, L. Carin, A general framework for reconstruction and classification from compressive measurements with side information, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2016, pp. 4239–4243.

[14] C.-G. Li, Z. Lin, J. Guo, Bases sorting: generalizing the concept of frequency for over-complete dicionaries, Neurocomputing 115 (2013) 192–200.

[15] S. Jiao, Design of optimal illumination patterns in single-pixel imaging using image dictionaries (2018). arXiv:1806.01340.
URL http://arxiv.org/abs/1806.01340

[16] Y. Li, C. Hedge, A. C. Sankaranarayanan, R. Baraniuk, K. F. Kelly, Compressive image acquisition and classification via secant projections, Journal of Optics 17 (2015) 065701–065710.

[17] R. P. W. Duin, E. Pekalska, The dissimilarity space: Bridging structural and statistical pattern recognition, Pattern Recognition Letters 33 (7) (2012) 826–832.

[18] A. Adler, M. Elad, M. Zibulevsky, Compressed learning: A deep neural network approach (2016). arXiv:1610.09615.
URL http://arxiv.org/abs/1610.09615

[19] A. C. Sankaranarayanan, M. A. Herman, P. Turaga, K. F. Kelly, Enhanced compressive imaging using model-based acquisition: Smarter sampling by incorporating domain knowledge, IEEE Signal Processing Magazine 33 (5) (2016) 81–94.

[20] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, 1st Edition, Springer Publishing Company, Incorporated, 2010.

[21] R. Rubinstein, OMP-Box v10 Matlab toolbox, last access: January 2019.
URL http://www.cs.technion.ac.il/~ronrubin/software.html

[22] M. Aharon, M. Elad, A. Bruckstein, K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation, IEEE Transactions on Signal Processing 54 (11) (2006) 4311–4322.

[23] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (1998) 2278–2324.

[24] The MNIST database of handwritten digits., last accessed: 14 February 2019.
URL http://yann.lecun.com/exdb/mnist/

[25] B. Marlin, K. Swersky, B. Chen, N. de Freitas, Inductive principles for restricted Boltzmann machine learning, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR, Vol. 9, 2010, pp. 509–516.

[26] The CalTech 101 silhouettes data set, last accessed: 14 February 2019.
URL https://people.cs.umass.edu/~marlin/data.shtml

[27] J. Xu, Q. Cai, S. C., DARC: Timely classification with randomly delayed features, in: Proceedings of the IEEE GLOBECOM Conference, 2016.