HOW TO READ FACES WITHOUT LOOKING AT THEM

A PREPRINT

Suyash Shandilya
Defence Research and Development Organisation
New Delhi - 110054
su.sh2396@gmail.com

Waris Quamer
Amdocs
Pune, Maharashtra - 411028
mr.warisquamer@gmail.com

April 24, 2020

ABSTRACT

Face reading is the most intuitive aspect of emotion recognition. Unfortunately, digital analysis of facial expression requires digitally recording personal faces. As emotional analysis is particularly required in more poised scenario, capturing faces becomes a gross violation of privacy. In this paper, we use the concept of *compressive analysis* introduced in [1] to conceptualise a system which compressively acquires faces in order to ascertain unusable reconstruction, while allowing for acceptable (and adjustable) accuracy in inference.

***K*eywords** Compressive analysis · Privacy · Emotion detection · Compressed sensing · Machine Learning

1 Introduction

Need for EQ As we move towards advancing artificial intelligence, we need to realise that intelligence is more than logical intellect. It is the ability to accurately perceive a data in a given context, and return an apposite response. Faster machines will render faster results, smarter machines will render smarter result. But the *hoi polloi* doesn't always seek an optical solution to a complex computational problem. Intuition entices more than intelligence. An 'IQ' alone cannot suffice to build human-like intuition. **Emotional Quotient** (EQ) is a major pillar for the foundation of future technology. Advances in deep learning have already shown good promise in the domain; what is required is an intent to accomplish it.

Need for Privacy Reading faces is one of the most trivial ways humans perceive each other's emotions. Our ability to express and read emotions via face and other physical cues is what advanced our species in communication. Other physiological signals like pulse rate, breathing pattern, also convey a more objective emotional analysis. EEG signals would probably top all of them but none of these can be as conveniently acquired as capturing facial expression. Nevertheless, all of these are privacy invasive. There is no prima facie way to infer something as personal as emotion, without invading some amount of privacy. The concept of *compressive analysis* allows one to seek a trade-of between privacy and personalisation.

1.1 Compressive analysis

Compressive Sensing [2, 3] is a recent idea which includes compression as a part of acquisition itself by acquiring random linear measurements instead of uniform samples. It further defines a lower bound on the number of compressed samples one needs to acquire to reconstruct the original data accurately; or to an accuracy limited only by the level of noise in acquisition. Interested readers are recommended to read the excellent introduction on the topic given by Candes and Wakin [4]. As the number of acquired measurements are reduced further, the reconstruction quality degrades to a level of unusability very soon. Even from such an irretrievable state, one can expect - given the art of compressive sampling - certain key structures to be preserved deep in the randomness of the compression. The idea of compressive analysis is to analyse this compressed representation (similar to a non-cryptographic digest) of the original image instead of the image itself. For a sufficient degree of compression, a reconstruction may not render a sensible image but

still be utilisable for certain important analysis (emotion detection, in this work). It maybe noted that since the analysis is performed only on the compressed version of the image, any reconstruction requires the knowledge of the sensing matrix ϕ which is (in our case) a binary gaussian random matrix of size $M \times N$ where M are the number of compressed samples and N is the length of the image vector (concatenated image matrix). Thus even if the reconstruction may render a 'sensitive' information in some sense, it can only be recovered by someone who has the knowledge of the matrix ϕ.

1.2 Previous work

This is a more advanced demonstration of the idea of compressive analysis first broached in [1]. As stated there, the concept of compressive classification was posited along with the single-pixel camera [5] by the name of smashed filters [**?**]. [6] demonstrates privacy preserving face recognition using secure multiparty computation. The idea in the paper was to have privacy preserving biometric verification system. Homomorphic encryption is a common idea in [6, 7] and many similar approaches to the idea. To the best of our research, this is the only paper so far, that assures privacy preserving emotion recognition from pure image analysis.

1.3 Paper Layout

The rest of the paper is covered as follows. Section 2 details the specifics of the methodology employed including the chosen dataset followed by the preprocessing (encryption/compression in our case) specifications. Section 3 presents the results of the chosen classification models and analyses their performances. Section 4 describes the reconstruction from the compressed images. Section 5 puts forward the conclusion and lays down the future scope of our research.

2 Methodology

2.1 Dataset

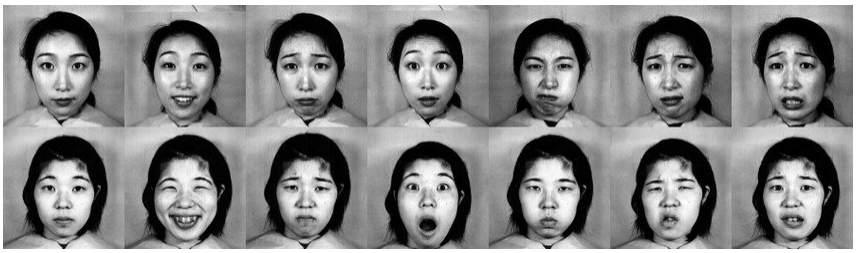

Figure 1: Samples from the Japanese Female Facial Expression (JAFFE) dataset

The Japanese Female Facial Expression (JAFFE) database contains 213 images of 7 facial expressions (6 basic facial expressions + 1 neutral) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects.

2.2 Image Encryption

Compressive Sensing requires that the vector being recovered should be sparse in some domain. The Daubechies-10 [8] wavelets were chosen for a sparse representation of the images here. The images underwent a 2D wavelet transformation before being concatenated to a column vector. As earlier in [1], binary gaussian matrix was chosen as a sensing matrix $\phi \in \{0,1\}^{M \times N}$ to simulate the DMD in the single-pixel camera [5]. For $M = 2000$, *some* reconstructions showed some characterisable facial features, thus it was decided that M must be less than 2000. For respectable classification accuracy, the lower bound for M was chosen as 50. To observe of the behaviour of the model for even lower values of M, the final values were chosen as: 800,500,200,100,50,20,10,5,2,1. $M = 50$ is equivalent to a compression ratio of 0.31% ($800 \equiv 5\%$).

2.3 Experimental Procedure

Each image from the dataset was read as 128x128 matrix. Fast Wavelet Transform (FWT) was then applied on the images obtaining the discrete wavelet coefficients at the maximum decomposition level (5 in this case) for the *db2* wavelet filter-bank. The coefficients were stacked to form the image vector of size $N \times 1$, which was multiplied with

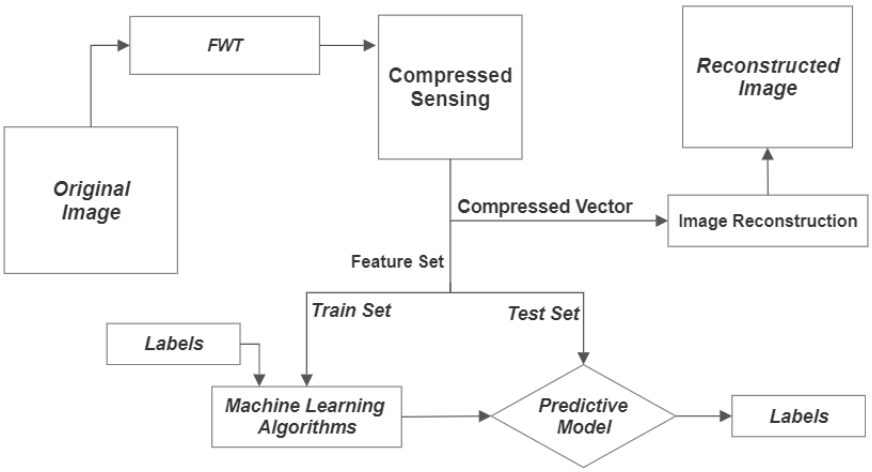

Figure 2: Systematic representation of our proposed work

the sensing matrix ϕ of size $M \times N$ (as mentioned in section 1.1 and 2.2) to obtain the compressed vector of the image. This compressed vector was used as the feature set for the classification models as well as the image reconstruction (see section 4). The systematic representation of our proposed work is explained in Figure 2.

3 Results and Analysis

For our analysis, the JAFFE dataset was classified into three broad emotional categories, as Positive, Neutral and Negative. Different classification models namely, Multilayer Perceptrons(MLPs), k-nearest neighbor, Decision tree, Support vector machines among others, were tested with various different parameter values. For evaluating the performance of the classifiers, a 5-fold cross validation accuracy score was used for the value of M = 500. The performances of different models for their optimised parameter values are shown in Table 1. With the ability to dynamically model non-linear and complex prediction functions, MLPs can learn hidden relationships without imposing any restrictions on the data. It is evident from the results that Multilayer Perceptron (MLP) was the best performing model whereas the rest of the classifiers had notably low accuracies as compared to that of the neural nets. Therefore, it was concluded that the MLPs have a better learning capacity and thus better suited for this problem. The performance of MLP was then further analysed for different values of M, the results of which are depicted in Table 2 and Figure 3. It can be observed that the accuracy of the model increases when the value of M is reduced from 800 to 500. One of the probable reasons for this behaviour could be a high correlation between attributes values for M = 800. For the lower values of **M** (i.e $M \in \{1, 2, 5\}$), the model seemed to be biased towards a particular class. There must have been too much loss in data for the model to be able to differentiate between classes.

Model	**Accuracy**
K-Nearest Neighbors	0.49
SVM (Linear kernal)	0.42
SVM (RBF kernal)	0.42
Gaussian Process	0.44
Decision Tree	0.28
Random Forest	0.26
Multilayer Perceptron	**0.79**
AdaBoost	0.44
Naive Bayes	0.30
QDA	0.30

Table 1: 5-Fold Cross Validation Score for Classifiers for **M = 500**

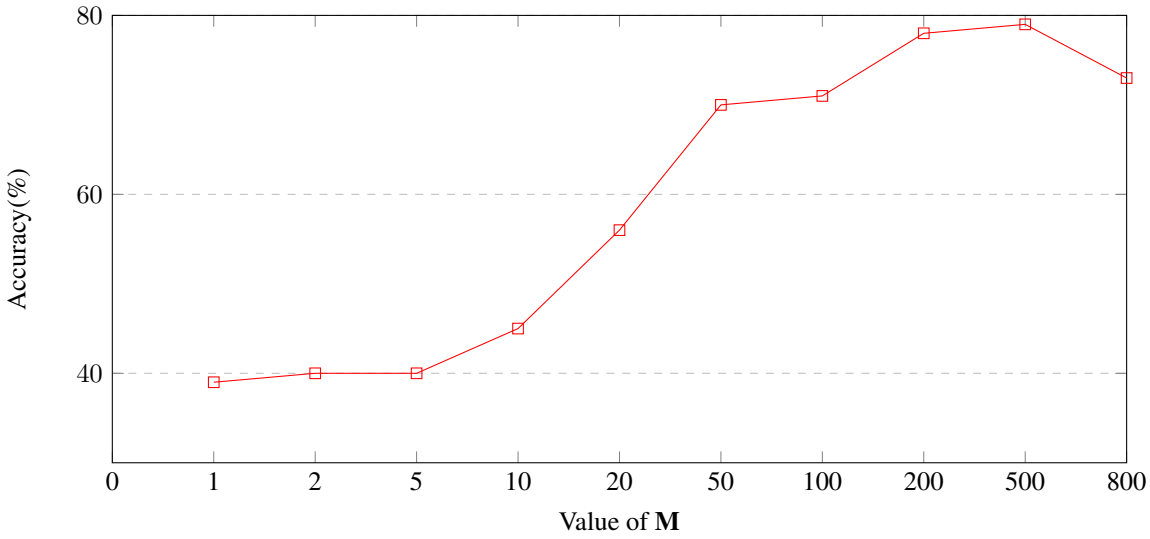

Figure 3: Accuracy of MLP Model Vs different Values of M

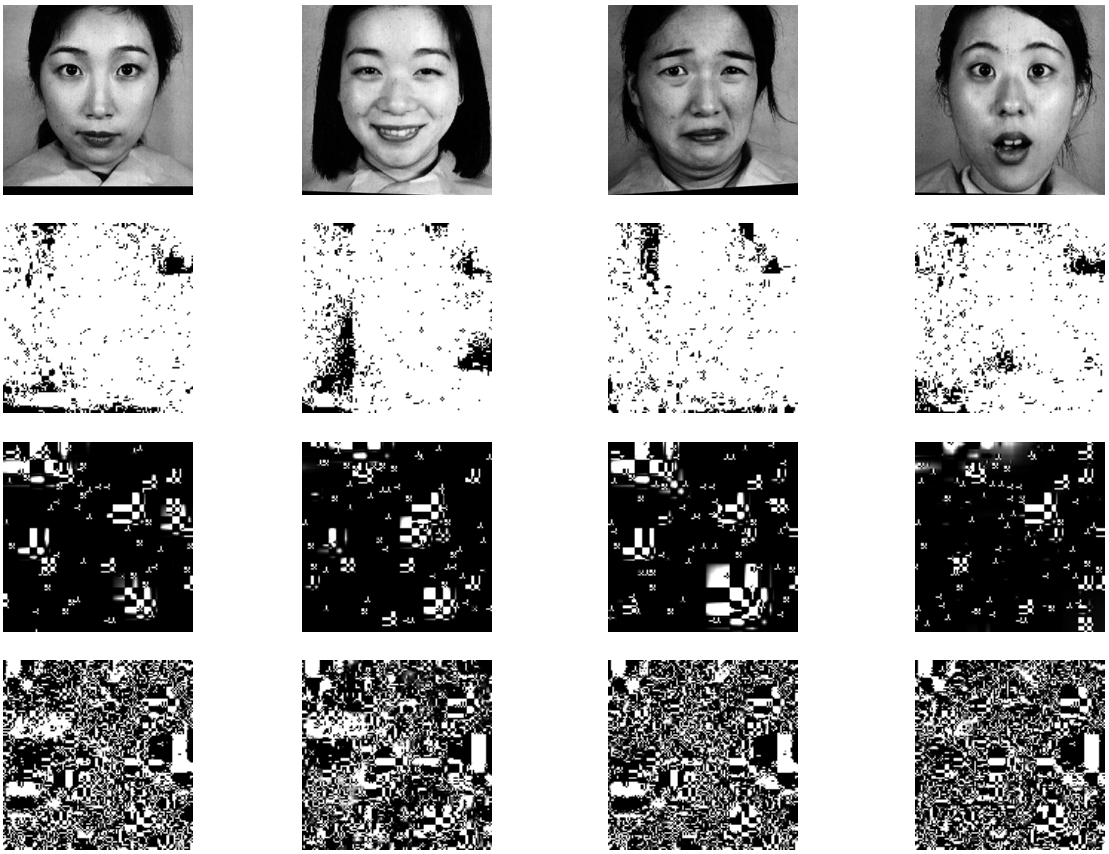

Figure 4: The first row consists of samples from our dataset for different labels (L to R - neutral, positive, negative, negative) and susequent rows are the reconstruction results for $M = 500, 50, 10$ respectively.

M	Compression Ratio(%)	Train Accuracy	Test Accuracy
800	4.89	0.90	0.73
500	**3.05**	**0.97**	**0.79**
200	1.22	0.96	0.78
100	0.61	0.95	0.71
50	0.31	0.91	0.70
20	0.12	0.74	0.56
10	0.06	0.62	0.50
5	0.03	0.52	0.40
2	0.01	0.43	0.40
1	0.006	0.43	0.39

Table 2: Accuracy of MultiLayer Perceptron for Different Values of **M**

4 Image Reconstruction

Typically, Matching Pursuit algorithms [9] are more popular in use because of the speed and ease in execution. Given our objective here is to ensure maximum security of the target data, we have stuck to using basis pursuit [10] for all our comparative reconstruction because of stronger guarantees of accuracy, albeit at the cost of computation capital. The SPGL1 [11] library was used to reconstruct the images. Since noise will only obfuscate reconstruction accuracy, and a well trained classifier can typically work as good in its presence, we did not consider any acquisition noise in our experiments. It was also noted that CVX [12, 13] gave slightly better reconstruction but was about 4 times slower. The images in Fig 4 may seem as a glitch to a passing reader but they are the best possible reconstruction we have been able to present. No single colormap could produce a more comprehensible image. One of the other shortcomings of Compressed sensing in general is that it will require very high bit depth to better portray the computation results. The checkered artefacts (more prominent in the bottom 2 rows) evince a very sparse vector in the wavelet basis. One may argue that better reconstruction is possible if we regularise the objective function with the l_2-norm. That may be true but it seemed practically impossible for us in our experiments. Moreover, the regularisation manifests more as noise than as clarity (if any). Nevertheless, there is a considerable room for improvement which seems more befitting to be mentioned in the next section.

5 Conclusion and Future Works

We have successfully demonstrated how - using a well trained classifier - extremely compressed samples which are unusable for a recognisable reconstruction, can be used to for classification even in subtle privacy sensitive applications like recognising facial expressions. Although only MLPs performed up to the mark in our analysis, other simple models can be reasonably effective if the degree of compression is reduced. The dataset used in our analysis has been made out of good and earnest efforts. The consistency of images makes computation easy. Unfortunately the real life scenario may be far from it. Our aim here was to prove the concept for face portraits. In future work, we will attempt to develop systems which are able to be at least as efficient on a more diverse and larger dataset. Significant improvement in reconstruction can be achieved by changing the basis of image representation to other more compatible bases. Intuitively it can be understood, how Ridgelets, or Curvelets [14] will be more compliant with portrait images than wavelets per se. More sophisticated splines can be developed to achieve an even better transformation. The intuition is to capture more structure of the image in the individual basis vectors. The same idea can be extended to the optimisation problem by utilising the block sparsity of the image vector in the wavelet basis. Visualising way ahead in time, if such a unit is deployed on a large scale, it could lead to having multiple different measurements of a person with the same expression. An unwanted adversary, who somehow manages to assimilate all these different measurements along with the sensing matrices may attempt to reconstruct certain faces. Even if a reconstruction seems very overreaching, a detection from a known subset of people could be feasible. In future we wish to analyse whether, and to what extent, such classifications are possible from compressed samples.

References

[1] Suyash Shandilya. Minimizing acquisition maximizing inference - a demonstration on print error detection. In *Algorithms for Intelligent Systems*. Springer, 2019.

[2] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, April 2006.

[3] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, Feb 2006.

[4] E. J. Candes and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, March 2008.

[5] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25(2):83–91, March 2008.

[6] Zekeriya Erkin, Martin Franz, Jorge Guajardo, Stefan Katzenbeisser, Inald Lagendijk, and Tomas Toft. Privacy-preserving face recognition. In Ian Goldberg and Mikhail J. Atallah, editors, *Privacy Enhancing Technologies*, pages 235–253, Berlin, Heidelberg, 2009. Springer Berlin Heidelberg.

[7] Ahmad-Reza Sadeghi, Thomas Schneider, and Immo Wehrenberg. Efficient privacy-preserving face recognition. In *International Conference on Information Security and Cryptology*, pages 229–244. Springer, 2009.

[8] Ingrid Daubechies. *Ten lectures on wavelets*, volume 61. Siam, 1992.

[9] Hongbo Bi, Chunhui Zhao, Ying Liu, and Ning Li. Performance evaluation of greedy reconstruction algorithms in compressed sensing. In *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1322–1327. IEEE, 2016.

[10] Alain Rakotomamonjy. Algorithms for multiple basis pursuit denoising. 2009.

[11] Ewout Van Den Berg and Michael P Friedlander. Probing the pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008.

[12] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

[13] Michael Grant and Stephen Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

[14] Jalal Fadili and Jean-Luc Starck. *Curvelets and Ridgelets*, pages 754–773. Springer New York, New York, NY, 2012.