Technical Report: models for the recognition of facial actions

Ing. Filippo Vella

Research founded by Short Term Mobility Program of National Research Council of Italy(CNR) and led at the Department of Computer Science of Boston University hosted by Prof. Stan Sclaroff

Introduction

Facial expressions and, in general, the movements expressed by face muscles are able to express a rich amount of information that convey a considerable part of communication. The focus of the present document is to survey techniques that can be adopted in a system focused on the extraction of information during interaction that are different from verbal communication. When verbal and audio information, that is principally involved in speech and verbal communication is discarded, the remaining part of the communication is conveyed with actions detected in the facial action and expressions. The shown techniques allow to detect face in a video stream and represent facial movements. The application of this knowledge can be applied to :

- affective computing
- intensional systems
- sign language

Affective Computing studies the systems that can recognize, classify and process human affects. It is focused on the emotions and an ideal system should be able to interpret the human behavior and give an appropriate response to the recognized emotions. Intensional systems are able to interpret the intension of a person in a context and can help him in a specific task. It is very helpful for person with disabilities[1]. Sign language systems are focused on the interpretation of the sign language that is mainly oriented to the movement of the hands and their postures. It is a complementary activity in the understanding of face movements during the dialogue with sign language since face expression and mouth movement can give additional and relevant information to the communication content.[2]. Some interesting technique to devise a system for expression understanding are described below. In the section 1 a very efficient algorithm for face detection is described. In section a commonly used standard for the description of face expression is shown. In section the representation of signals with Compressive Sensing techniques is motivated. Some work have used Compressive Sensing for face expression recognition and an analysis id done in 3.1

1 Face Detection

There are many techniques to detect faces in still images and video sequences but the most used adopted in consumer and scientific field is the algorithm proposed by Viola and Jones.[3] . The method is based on the algorithm called Ada Boost [4] and allows to create a strong classifier starting from a set of weak ones. For the particular application dealing with face detection a set of simple filters as Haar filters are used. The Haar filters are shown in figure 1 and are easy to compute and

some hint about how to calculate these filters in a quick way adopting the integral image is given in [3].

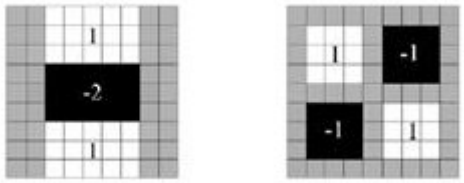

Figure 1: Haar Filter Examples

The methodology, in general has as input a training set, $\mathbf{X} = (x_1, y_1), \cdots, (x_n, y_n)$ where x_i is a representation of the generic sample and y_i indicates the presence or the absence of the specific item of interest(a face in this case).
The weak classifiers are named h_j and can detect if the chosen object is present in the given sample with a performance that is better (also slightly better) than the random guessing. It can be also said that the classifier is slightly correlated with the true classification. From the collection of a set of weak classifiers can be created a new composite classifier composed by a weighted sum of the classifiers. The new classifier is a strong classifier that is tightly correlated with the true classification and can reliably detect the presence of a given signal. The algorithm to set the weights is shown in algorithm 1 at page 3:

Algorithm 1 Ada Boost

Initialize weights $w_{1,i}$
for $t = 1, \cdots, T$ **do**
normalize weights $w_{t,i}$
evaluate the weighted error for each weak classifier
$\epsilon_j = \sum_i w_i \left| h_j(x_i) - y_i \right|$
choose the h_t with the lowest ϵ_t
update the weights:
$w_{t+1,i} = w_{t,i} \cdot \beta_t^{1-e_i}$
where $e_i = 0$ if x_i is classified correctly, $e_i = 1$ otherwise
$\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$
end for
Final Classifier:

$$H_{final}(x) = \begin{cases} 1 & if \quad \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & otherwise \end{cases} \tag{1}$$

where $\alpha_t = log \frac{1}{\beta_t}$

At each step the correct of wrong answer of the $j-th$ weak filter is calculated for each $i-th$ training sample. The response of the $j-th$ is estimated through :

$$\epsilon_j = \sum_i w_i \left| h_j(x_i) - y_i \right| \tag{2}$$

The best classifier, among the weak ones is the classifiers that scores the lowest error. The weights w_i play an important role in the selection of this classifier taking into account the most important samples. At the first iteration the weights are normalized so that their sum produce a given value(e.g. one) and all of them have the same value. This normalization assures that all the samples have the same weight and the winning classifiers is the one that produces the highest number of correct responses on the training set (that is equivalent to the least number of wrong classification). The best weak classifier is called h_t and will classify some samples in a correct way and some other in a wrong way. The weights are updated according to equation (3)

$$w_{t+1,i} = w_{t,i} \cdot \beta_t^{1-e_i} \tag{3}$$

where e_i is equal to 0 if the $i-th$ sample is correctly classified and is equal to 1 if the sample is wrongly classified. If a sample is wrongly classified the value of the weight remains unchanged. This update brings that if the $i-th$ sample is correcly classified the weight is multiplied for β_t. Where

$$\beta_t = \frac{\epsilon_t}{1-\epsilon_t} \tag{4}$$

Since ϵ_t is the error produced by the chosen h_t classifier, the better is the performance of the classifier, the lower is the value of ϵ_t. If the value of ϵ_t is low, consequently also β_t is low and the weight for the correctly classified samples is reduced making them less important.

2 Facial Action Recognition

Facial Action Coding System (FACS)[5] is a standard technique for describing expression and activities of human face. There is no information of motion but a face is represented statically according 44 Action Units (AU). Many techniques and models have been proposed to recognize the single Action Units present in a image. Among the most promising techniques for the recognition of action units with temporal information there are the works of Tong et al. [6] and Simon et al. [7]. A general overview on the approaches for AU recognition is given in [8].

Tong et al. in their work [6] propose a unified probabilistic framework based on Dynamic Bayesian Network(DBN) to represent face motions. Simon et al. in [7] developed a segment based approach that employs Bayesian Network with Support Vector Machine(SVM) to set a robust dynamic recognition in time sequences about

face movements. In [9] an analysis of the most used dataset (Cohn-Kanade) for Action Unit Recognition has been done. [1]. A description of the dataset CK and CK+ made by the author of the same dataset is given in [10] [11].

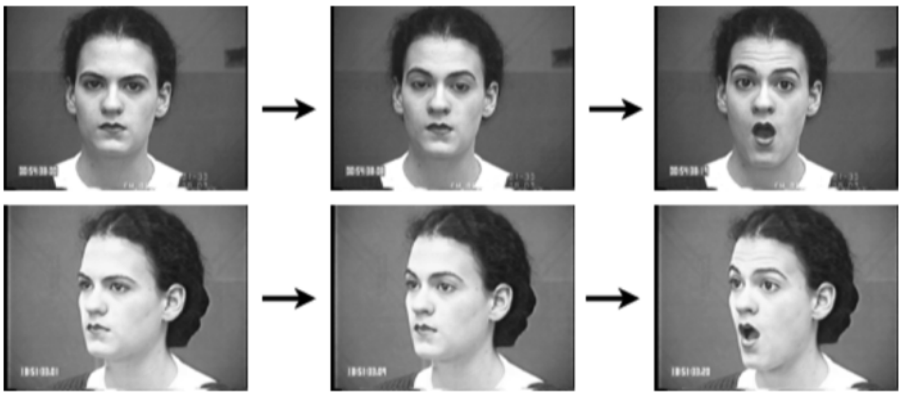

Figure 2: Examples of Face Actions : Frontal and 30-degree views from the Cohn-Kanade database. Each sequence begins with a neutral expression and proceeds to a target expression

In particular the combinations of Face Actions have been analyzed by Mahoor et al. in [9] and the combination with highest frequency have been selected. A sub set of them are: AU 1+2+5+27, 15+17, 6+12+25, 4+9+17+23, 20+25. An extenstive description of these facial actions with a relationship with facial muscles that activate particular expressions can be found in [10].

3 Sparse Representation

The possibility to adopt sparse representation has grown in the last years due to the new results provided in compressing sampling theory. Conventional approaches for image acquisition and in general for signals recording is based on the basic principle of the Nyquist frequency sampling theory.

The emerging theory of Compressing Sensing [12] [13] states that is possible to capture signals, and therefore images, with a sensibly reduced number of samples if a slight corruption is allowed. The Compress Sensing Theory is tightly coupled with Sparse Representation since a signal can be represented through a linear combination of few non zero coefficients.The condition is that the representation is made on a redundant basis that constitutes an overcomplete dictionary. An introductory presentation of this theory is given in [14]. The Compressive Sensing is focused on the reduction of the number of measurements that are stated in the sampling theory. This reduction is achieved exploiting the *compressibility* of signals. The focal point

[1]The datased is available at `http://vasc.ri.cmu.edu/idb/html/face/facial_expression`

(a) P 10 : AU 1 + 2 + 5 + 25 + 27

(b) P 14 : AU 4 + 7 + 10 + 17 + 23 + 24 + 43

(c) P 35 : AU 6 + 12 + 25

(d) P 42 : AU 1 + 2 + 5 +25 + 27

(e) P 56 : AU 15 + 17

(f) P 138 : AU 25 + 27

Figure 3: Examples of most frequent combination of Action Units

is that the measurement is not made with simple sampling (evaluating the signal in a point) but are measured through a function. If we consider a real-valued, finite-lenght, one-dimensional, discrete-time signal x, it can be seen as a column of N values in $\mathbb{R}^N$ space with $n = 1, 2, 3, \ldots, N$. Any signal can in $\mathbb{R}^N$ can be represented according to a basis of vectors $\in \mathbb{R}^N : \{\psi_i\}_{i=1}^N$. Considering that these vectors are orthonormal they form a basis matrix $\Psi := [\psi_1|\psi_2|\ldots|\psi_N]$. For the definition of a basis, any signal with dimension N can be expressed as :

$$\mathbf{x} = \sum_{i=1}^{N} s_i \psi_i \quad \text{or} \quad x = \Psi \mathbf{s} \tag{5}$$

where $\mathbf{s}$ is the Nx1 column vector of weighting coefficients $s_i = \langle \mathbf{x}, \psi_\mathbf{i} \rangle = {\psi_\mathbf{i}}^T \mathbf{x}$, where $\cdot^T$ denotes the Hermitian transpose operation. The representation $\mathbf{x}$ in the time domain is equivalent to $\mathbf{s}$ in the Ψ domain.

Compressive Sensing is focused on signals that have a sparse representation and can be represented with the linear combination of K basis vector with $K \ll N$. The equivalent is that only K value of equation (5) are non zero while the other $N - K$ are zero. Sparsity is motivated by the fact that many signals are compressible and there exist a basis Ψ where the representation has few large coefficient and many small coefficients. These signals are usually represented with few coefficient

employing the *transform coding*. This technique is used in audio and video standards to compress natural signals. For JPEG image compression format the discrete cosine transform (DCT) is adopted and, in an analog way, a decomposition on wavelets is used for JPEG-2000 compression algorithm. A considerable part of the energy of the signal is concentrated on few coefficient while most of the coefficient are small and can be discarded (allowing a non-perfect reconstruction). The state-of-art algorithms for the representation of signals and multimedia items are deviced in two phases: a sampling phase and a compression phase. The sampling is done according the Nyquist theory acquiring the full N-sample signal $\mathbf{x}$. The coefficients s_i are computed calculating $s = \Psi^T \mathbf{x}$ and that retaining the largest K value. The $N - K$ values are discarded. At last the K values are compressed with an entropy coding. The used scheme has different drawbacks:

- a potentially large number of sample must be considered even if the value of K is small
- the encoder must compute N coefficient s_i although only K are retained
- the largest coefficient must be located

The alternative process based on compressive sensing tends to acquire a signal through a compressed representation without sampling N values. For this process a set of M linear mesurements are used. The measurement is achieved applying the inner product between the signal $\mathbf{x}$ and a collection of M vectors $\{\phi_j\}_{j=1}^{M}$, producing the values $y_j = \langle \mathbf{x}, \phi_j \rangle$. Packing the y_j values in the vector $\mathbf{y}$ and the vectors ϕ_j^T as rows in the matrix MxN matrix Φ, from equation (5) can be written:

$$\mathbf{y} = \Phi \mathbf{x} = \Phi \Psi \mathbf{s} = \Theta \mathbf{s} \tag{6}$$

where Θ is a MxN matrix given by the product of $\Theta := \Phi\Psi$. A graphical representation of the process is shown in figure 4

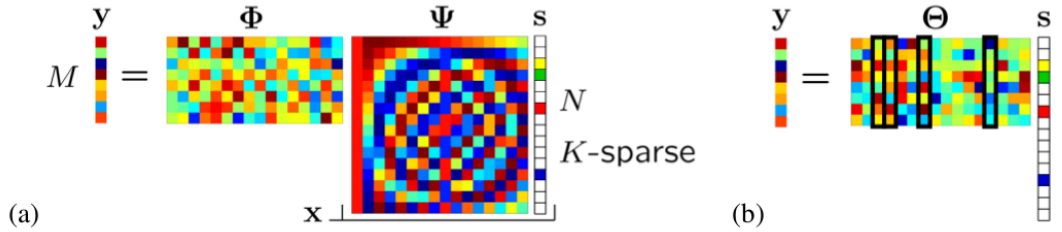

Figure 4: Compressive Sensing Measurement Process

It can be noticed that the measurement process is not adaptive since the matrix Φ is not dependent from the data. What is done in Compressive Sensing is to design the matrix Ψ (called Measurement Matrix) so that the signal can be reconstructed

from a limited number of measurements (M) and to approximate the number of non-zero values of the vector $\mathbf{s}$. The technique of the compressive sensing can be divided in two parts: the generation of the a stable measurement matrix Ψ and the reconstruction of the signal.

3.0.1 Stable Measurement Matrix

The measurement matrix Ψ must be chosen so that it assures that the salient information in any K-sparse signal is not damaged by the dimensionality reduction from $\mathbf{x} \in \mathbb{R}^N$ to $\mathbf{y} \in \mathbb{R}^M$. The target of the technique is to reconstruct the $\mathbf{x}$ signal (or in an equivalent way the vector $\mathbf{s}$ in the basis Ψ) starting from the M measurement of the $\mathbf{y}$ vector.

The measurement should not destroy the information in $\mathbf{x}$. In general the equation (6) asserts that the $\mathbf{s}$ can be obtained by $\mathbf{y}$ with the solution of the linear algebra system. In this case, since $M < N$, the problem is ill posed since there are fewer equation that unknowns.

A solution could be easy to find if, considering that the $\mathbf{s}$ vector is sparse, the values that are different from zero in the sparese vector $\mathbf{s}$ would be known. If the highlighted values shown in figure 4 (b) would be the only values different from zero it would be sufficient to select the column of matrix Θ to obtain a linear system with M equations and K unknowns that is well-conditioned (if M $\geq K$) and a stable inverse matrix can be calculated. A necessary and sufficient condition so that the MxK system is well-conditioned is that for any vector $\mathbf{v}$ sharing the same K non zero entries with $\mathbf{s}$ holds the property:

$$1 - \epsilon \leq \frac{\left\|\mathbf{\Theta v}\right\|_2}{\left\|\mathbf{v}\right\|_2} \leq 1 + \epsilon \tag{7}$$

for some $\epsilon > 0$. The equation (7) is equivalent to state the the matrix Θ preserve the lengths of K-sparse vectors.

In the operative application of this technique it is unknown the number of non zero values and the locations of the $\mathbf{s}$ signal. In this case is used the Restricted Isometry Property (RIP) [13] that states the a stable inverse for K-sparse and compressible signal is available if Θ satisfy property 7 for an arbitrary $3K$-sparse vector $\mathbf{v}$.

An alternative approach to stability is to create a measurement matrix Φ that is incoherent with the sparsifying matrix Ψ. In other words the vectors $\{\phi_j\}$ cannot represent sparsely the $\{\psi_i\}$ and viceversa[13][12]. The behaviour is similar to the representation of Fourier representation with delta spikes and sinusoids.

Given a matrix Ψ the choice of a Φ matrix should be selected so that the matrix Θ verify the Restricted Isometry Property. Given the matrix Θ, that is calculated from $\Psi\Phi$, the check that it satisfies the RIP is combinatorically complex as $\binom{N}{K}$.

To avoid this calculation the matrix Φ is filled with random values that are indipendently and identically distributed (i.i.d) from a zero mean, $1/N$-variance Gaussian density. Then the measurement of the $\mathbf{y}$ is done considering M differently weighted linear combinations of the element of $\mathbf{x}$. The choice of a Gaussian Φ has some important properties:

- The matrix Φ is incoherent with the basis $\Psi = \mathbf{I}$ of delta spikes with high probability. It would take N spikes to represent each row of Φ.
- Due to the properties of the iid Gaussian distribution used to generate Φ, the matrix $\Theta = \Phi\Psi$ is also iid gaussian regardless the choice of the sparsifying matrix Ψ. In other words, the random Gaussian measurements Φ are universal in the sense that Θ matrix has the RIP property with high probability for every possible Ψ. An alternative technique is to use random matrices with random ± 1 entries proposed by Rademacher [15] that also show to have RIP and universality property.

3.0.2 Signal Reconstruction Algorithm

The Isometry Property (RIP) provides the proof that a K-sparse signal can be described with M measurement in $\mathbf{y}$. The reconstruction phase must take the measurement in $\mathbf{y}$, the random measurement matrix Φ and the sparsifying basis Ψ and generate the length-N signal $\mathbf{x}$ or in an equivalent way the sparse coefficient vector $\mathbf{s}$.

Since the vector $\mathbf{y}$ has M values and $M < N$ there are infinitely many $\mathbf{s}'$ that satisfy:

$$\Theta\mathbf{s}' = \mathbf{y} \tag{8}$$

all these vectors lie on the $(N - M)$-dimensional hyperplane $\mathcal{H} := \mathcal{N}(\Theta) + s$ in $\mathbb{R}^N$ corresponding to the null space $\mathcal{N}(\Theta)$ of Θ translated to the true sparse solution $\mathbf{s}$.

If $\Theta\mathbf{s} = \mathbf{y}$ then $\Theta(\mathbf{s} + \mathbf{r}) = \mathbf{y}$ for any vector $\mathbf{r}$ in the null space.

The goal is to find the signal's sparse coefficient vector $\mathbf{s}$ in the translated null space. Generalizing the norm l_p of a vector $\mathbf{s}$ as :

$$\left\|\mathbf{s}\right\|_p = \left[\sum_{i=1}^{N} |s_i|^p\right]^{1/p} \tag{9}$$

Varying of the parameter p is possible to optimize different norms:

- l_2 : the classical approach is to consider the least squares criteria that is it is searched the vector in the translated nullspace $\mathcal{H}$ with the smallest l_2 norm:

$$\hat{s} = \text{argmin}\left\|s'\right\|_2 \text{ such that } \Theta\mathbf{s}' = \mathbf{y} \tag{10}$$

There is a closed form solution $\hat{s} = \Theta^T(\Theta\Theta^T)^{-1}\mathbf{y}$ that allow to recover the value of $\hat{s}$ but the solution found, instead of being a K-sparse approximation, is an approximation of $\mathbf{s}$ with a plenty of ringing.

- l_0 : an alternative to l_2 norm is to consider the l_0 norm. The searched vector is the sparsest vector (that is the vector with the highest number of zero values) in the translated null space $\mathcal{H}$:

$$\hat{s} = \text{argmin}\left\|s'\right\|_0 \text{ such that } \Theta\mathbf{s}' = \mathbf{y} \tag{11}$$

  Unfortunately the solution of (11) is numerically unstable and requires to compute a NP-complete problem with all the combination for the non zero value possible position in $\mathbf{s}$ (that are $\binom{N}{K}$)

- l_1 is the norm used in compressive sensing since for $\mathbf{y}$ vector with size $M \geq cKlog(N/K)$ measured with Gaussian process can reconstruct K-sparse vectors with high probability [13][12]

$$\hat{s} = \text{argmin}\left\|s'\right\|_1 \text{ such that } \Theta\mathbf{s}' = \mathbf{y} \tag{12}$$

  The process of searching the $\hat{s}$ is a convex optimization problem that can be solved with *basis pursuit* technique with a computational complexity of $O(N^3)$[13][12]

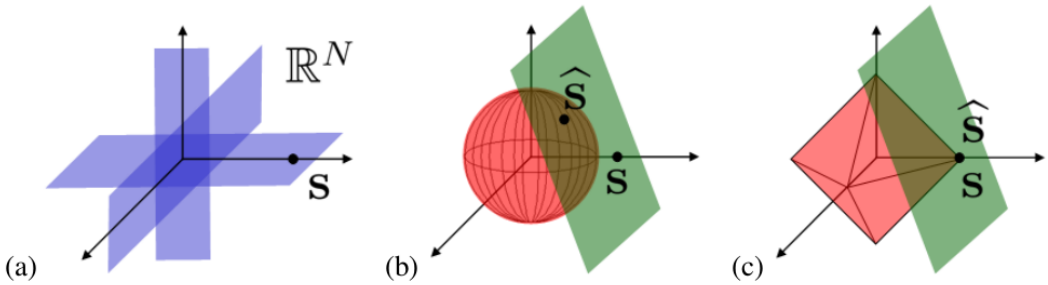

Figure 5: Search for the approximation of $\mathbf{s}$ signal with l_0(a), l_1(b) and l_2(c) norm

In the figure 5 is shown a geometric interpretation of the results with different norms. For the definition of sparse vectors, the set of K-sparse vectors $\mathbf{s}$ in $\mathbb{R}^N$ is a highly non linear space consisting of all K-dimensional hyperplanes that are aligned with coordinate axes. Sparse vectors are placed close to the coordinate axes in $\mathbb{R}^N$. In figure 5(b) shows that the translated null space $\mathcal{H} = \mathcal{N}(\Theta) + \mathbf{s}$ is a hyperplane of dimension $(N - M)$ and is oriented at a random angle due to the randomness of matrix Θ. In the figure, a sketch in three dimension is proposed while typical problems have dimensions much higher. The l_2 minimizer $\hat{s}$ from equation (10) is the

point on $\mathcal{H}$ that is closest to the origin. It can be seen as a point on a hypersphere that is tangent with the $\mathcal{H}$ hyperplane.

Since the orientation of $\mathcal{H}$ is random the closest point $\hat{s}$ will probably be in a different point from $\mathbf{s}$ and will be neither sparse nor a good approximation for the point $\mathbf{s}$. In Figure 5 (c) is shown the equivalent of the hypersphere with norm l_1. In this case the polytope will be aligned with the coordinate axes (and it becomes pointier as the dimension N grows) and considering a growing dimension l_1 hypersphere, it will touch the translated null space $\mathcal{H}$ at a point near the coordinates and precisely where the sparse vector $\mathbf{s}$ is located.

3.1 Compressive Sensing for face activity classification

This theory of Compressive Sensing has been adopted for face recognition and for face expression recognition by Wright et al. [16], Ying et al. [17], and by Mahoor et al. [9].

Wright et al. [16] attack the problem of face recognition through a linear regression model and exploit the theory of Compressive Sensing to classify unknown faces. A single face is represented as a combination of multiple given faces with known identities. (This set of faces forms the overcomplete dictionary). The employment of the Compressive Sensing Theory is motivated by the authors as they state that face recognition with occlusions and corruption is a task intrinsically sparse with comparison to pixel level.

Ying et al. [17] adopted a sparse representation for facial expression recognition. Two classifiers are used to discriminate the sparse values: raw gray scale pixel and local binary patterns. The results were obtained fusing the results from the two classifiers. Mahoor et al. in [9] state that the authors did not justify the design of the used over complete dictionary and the motivation for the application of the L_1 norm. The problem is not trivial and the question is when the Compressive Sensing can be applied ? Which are the condition for the application of the L_1 minimization? Mahoor et al [9] apply the Compressive Sensing to faces when they are represented by Action Units (AU) described by Facial Action Coding System (FACS)[5].

4 Conclusions

Pattern analysis and retrieval methods for automated labeling of humans and analysis of facial motions in unconstrained digital video sources is a very interesting topic that can bring to a large amount of applications in multiple aspects of human machine interaction. Action classification that aims at expanding action classification to include head and facial movements which can be related to people expressions in video clips. This also included examinations of issues related to analysis of facial and head gestures during sign language communication, detect people emotions and in general their intentions.

References

[1] Infantino I., Lodato C., Lopes S., and Vella F., "Implementation of a intentional vision system to support cognitive architecture," in *Proceedings of 3rd International Conference on Computer Vision Theory and Applications - VISAPP International Workshop on Robotic Perception (VISAPP-RoboPerc08)*, 2008.

[2] Ashwin Thangali, Joan P. Nash, Stan Sclaroff, and Carol Neidle, "Exploiting phonological constraints for handshape inference in asl video," in *CVPR*, 2011, pp. 521–528.

[3] Paul Viola and Michael Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[4] Yoav Freund and Robert E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *European Conference on Computational Learning Theory*, 1995, pp. 23–37.

[5] P. Ekman and W. Friesen, *Facial Action Coding System: Manual*, Consulting Psychology Press, 1978, Palo Alto.

[6] Yan Tong, Jixu Chen, and Qiang Ji, "A unified probabilistic framework for spontaneous facial action modeling and understanding," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, pp. 258–273, February 2010.

[7] T. Simon, M.H. Nguyen, F. de la Torre, and J.F. Cohn, "Action unit detection with segment-based svms," in *CVPR10*, 2010, pp. 2737–2744.

[8] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 1, pp. 39–58, January 2009.

[9] Mohammad H. Mahoor, Mu Zhou, Kevin L. Veon, S. Mohammad Mavadati, and Jeffrey F. Cohn, "Facial action unit recognition with sparse representation," in *Proc. of IEEE Automatic Face and Gesture Recognition*, 2011.

[10] T. Kanade, J. F. Cohn, and Yingli Tian, "Comprehensive database for facial expression analysis," *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, pp. 46–53, 2000.

[11] P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *CVPR4HB10*, 2010, pp. 94–101.

[12] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transaction on Information Theory*, vol. 52(2), 2006.

[13] D. Donoho, "Compressed sensing," *IEEE Transaction on Information Theory*, vol. 52(4), 2006.

[14] Richard Baraniuk, "Compressive sensing," *Lecture Notes in IEEE Signal Processing Magazine*, vol. 24, 2007.

[15] Richard Baraniuk, Mark Davenport, Ronald Devore, and Michael Wakin, "The Johnson-Lindenstrauss lemma meets compressed sensing," http://dsp.rice.edu/cs/jlcs-v03.pdf.

[16] John Wright, Allen Yang, Arvind Ganesh, Shankar Sastry, and Yi Ma, "Robust face recognition via sparse representation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 31(2), 2009.

[17] Zi-Lu Ying, Zhe-Wei Wang, Ming-Wei Huang, De-Shuang Huang, Xiang Zhang, Carlos Reyes Garcia, and Lei Zhang, "Facial expression recognition based on fusion of sparse representation," in *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence -LNCS*, 2010, vol. 6216.