

# POLYNOMIAL BASED TEXTURE REPRESENTATION FOR FACIAL EXPRESSION RECOGNITION

Cristina Bordei<sup>\*</sup>, Pascal Bourdon<sup>†</sup>, Bertrand Augereau<sup>†</sup>, Philippe Carré<sup>†</sup>

<sup>\*</sup> Technicolor, 975, avenue des Champs Blancs, 35576 Cesson-Sevigné, France  
e-mail: firstname.lastname@technicolor.com

<sup>†</sup> XLIM-SIC Laboratory, UMR CNRS 7252 Boulevard Marie et Pierre Curie, 86962 Chasseneuil France  
e-mail: firstname.lastname@univ-poitiers.fr

## ABSTRACT

In this paper, we propose a new polynomial based texture representation method for extracting information about facial expressions. While many appearance-based methods have been proposed over the years to improve the performance of facial expression recognition, most descriptors are usually unable to both provide precise multi-scale / multi-orientation analysis and handle the redundancy problem effectively.

We will explain how coefficients obtained from polynomial projections of pixel intensities on a complete basis can be used for compact, hierarchical image approximation and structural analysis. We have tested our approach on two publicly available databases and achieved encouraging results comparable to the state of the art.

**Index Terms**— Complete Polynomial Basis, Facial expressions, Appearance-based classification

## 1. INTRODUCTION

The automatic recognition of facial expressions is one of the most challenging and popular topics in the computer vision domain as it impacts important applications such as virtual reality, broadcasting, user profiling or video conferencing.

An essential step for a successful facial expression recognition is the extraction of facial features that attempt to find the most effective representation of face images. There are two common feature extraction approaches : geometric feature-based systems, using major face components and/or feature points, and appearance based systems using image filters. A thorough survey of the existing work can be found in [1, 2, 3]. Experimental results show that methods using Gabor wavelet transforms, derived from biological principles on the visual system, provide superior performance and are an effective method for facial expression recognition [4, 5]. However, it is both time and memory intensive to convolve face images with a bank of Gabor filters to extract multi-scale and multi-orientation coefficients.

Polynomial representations are similar to complete wavelet packet decompositions for a defined scale. Such descriptions have been used for the characterization and representation of handwritten mathematical symbols [6], the analysis of vowels and consonants in spectral frequency for speech recognition [7], or the generation of linear phase two-dimensional FIR digital filter functions [8]. Their use in image representation has also been demonstrated in [9, 10].

In [11], Carré and Augereau proposed a multi-scale hypercomplex 2D polynomial transform for color images based on quaternionic polynomials.

In this paper, we investigate the use of coefficients resulting from polynomial projections for texture representation within a system of facial expression recognition. Details about polynomial texture representation are provided in section 2, while polynomial decomposition is presented in section 3. We show in section 4 our proposed method and we compare in term of computational efficiency polynomial transforms to Gabor transforms. Experimental results obtained by applying the proposed technique on MUG [12] and the extended Cohn-Kanade [13] databases are provided in section 5 and show significant recognition rates over state-of-the-art methods. Finally, we will conclude and open the discussion on further works in section 6.

## 2. COMPLETE BASES

Our motivation to use a polynomial representation is that the orthogonal polynomials have, similarly to Gabor filters, some properties related to the human visual system [14], including multi-scale / multi-resolution information representation.

Let a Real Bivariate Polynomial of degree  $d$  be the function of  $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{R}^2$  defined as :

$$P(\mathbf{x}) = \sum_{\substack{(d_1, d_2) \in [0; d]^2 \\ d_1 + d_2 \leq d}} a_{d_1, d_2} x_1^{d_1} x_2^{d_2} \quad (1)$$

where  $d_1 \in \mathbb{N}^+$  and  $d_2 \in \mathbb{N}^+$  are the degrees of variables  $x_1, x_2$  and the  $\{a_{d_1, d_2}\} \in \mathbb{R}$  are the coefficients of the po-

polynomials. The overall degree of the polynomials is then the maximum of  $d_1 + d_2$ .

Considering a finite set of pairs  $D = \{(d_1, d_2)\} \subset \mathbb{N}^2$ , we represent by  $\mathbb{E}_D$  the space of all real bivariate polynomials such as  $a_{d_1, d_2} \equiv 0$  if  $((d_1, d_2) \notin D)$  and by  $\mathcal{K}_D$  the subset of real monomials :

$$\mathcal{K}_D = \left\{ K_{d_1, d_2}(\mathbf{x}) = x_1^{d_1} x_2^{d_2} \right\}_{(d_1, d_2) \in D} \quad (2)$$

Obviously  $\mathcal{K}_D$  satisfies the linear independence and spanning conditions and so,  $\mathcal{K}_D$  is a basis of  $\mathbb{E}_D$ , the canonical basis. In image analysis, we look for bases with suitable properties such as orthogonality or normality. So, to construct a discrete orthonormal real bivariate polynomial finite basis we first have to consider the underlying discrete domain :

$$\Omega = \left\{ \mathbf{x}_{(u,v)} = (x_{1,(u,v)}, x_{2,(u,v)}) \right\}_{(u,v) \in D_1} \quad (3)$$

where  $D_1$  represents the set of pairs associated to  $\Omega$ .

Starting from  $\mathcal{K}_D$  we intend to construct a new orthonormal basis by applying the Gram-Schmidt process. That implies that we need some product and norm for real bivariate functions defined on  $\Omega$ . Taking into account the computational contingencies, given two real bivariate functions,  $F$  and  $G$ , their discrete extended inner product is defined by :

$$\langle F | G \rangle_w = \sum_{(u,v) \in D_1} F(\mathbf{x}_{(u,v)}) G(\mathbf{x}_{(u,v)}) w(\mathbf{x}_{(u,v)}) \quad (4)$$

with  $w$  a real positive function over  $\Omega$  (Legendre, Chebichev, Hermite, ...). Then, the actual construction process of an orthonormal basis :

$$B_{D_1, w} = \{B_{d_1, d_2}\}_{(d_1, d_2) \in D_1} \quad (5)$$

is a recurrence upon  $(d_1, d_2)$  :

$$T_{d_1, d_2} = K_{d_1, d_2} - \sum_{(l_1, l_2) \prec (d_1, d_2)} \langle K_{d_1, d_2} | B_{l_1, l_2} \rangle_w B_{l_1, l_2} \quad (6)$$

$$B_{d_1, d_2}(x) = \frac{T_{d_1, d_2}}{|T_{d_1, d_2}|_w} \quad (7)$$

where  $\prec$  is the lexicographical order and  $|\cdot|_w$  the norm induced by  $\langle \cdot | \cdot \rangle_w$ . The resulting set of  $B$  polynomials verifies :

$$\langle B_{d_1, d_2} | B_{l_1, l_2} \rangle_w = \begin{cases} 0 & \text{if } (d_1, d_2) \neq (l_1, l_2) \\ 1 & \text{if } (d_1, d_2) = (l_1, l_2) \end{cases} \quad (8)$$

$B_{D_1, w}$  is effectively an orthonormal basis with respect to a weighting function  $w$ . A special case is the *complete basis* where  $D_1$  represents exactly the set of pairs associated to  $\Omega$ , that is

$$D_1 = [0; N_1] \times [0; N_2] \quad (9)$$

A complete basis, related to the discrete extended inner product (4) is the orthonormal basis whose domain is  $\Omega$  defined by the family :

$$\{B_{d_1, d_2}(\mathbf{x})\}_{\substack{d_1=0..n_1 \\ d_2=0..n_2}} \quad (10)$$

The number of polynomials in the complete polynomial basis is given by the size  $(n_1 + 1) \times (n_2 + 1)$ .

### 3. POLYNOMIAL DECOMPOSITION

To acquire an efficient strategy in image analysis, we need a joint spatial/frequency representation. In this section, we show that real discrete orthonormal polynomials can be considered as a discrete multiscale decomposition.

Considering a function  $U$  defined on a domain  $\Omega$  of  $n_1 \times n_2$  sizes and a basis of  $h_1 \times h_2$  sizes, the decomposition process is expressed, at a step  $L$ , according to :

1. partition of the discrete domain  $\Omega^L$  with a number of  $\Delta$  sublattices, of sizes  $h_1^L \times h_2^L$  ;
2. for each subinterval  $\Delta$ , approximation of the corresponding restriction  $U^L$  in a complete basis constructed on  $\Delta$ . The polynomials coefficients are defined as :

$$b_{i,j}(U^L) = \langle U^L | B_{i,j} \rangle_w \quad (11)$$

3. the reordering or orthogonal polynomial coefficients  $b$  into  $h_1^L \times h_2^L$  functions  $U_{i,j}^{L+1}$ , on domains of

$$\left[ n_1^{L+1} \equiv \frac{n_1^L}{h_1^L} \right] \times \left[ n_2^{L+1} \equiv \frac{n_2^L}{h_2^L} \right] \quad (12)$$

sizes to provide image subbands in a multiresolution decomposition-like structure.

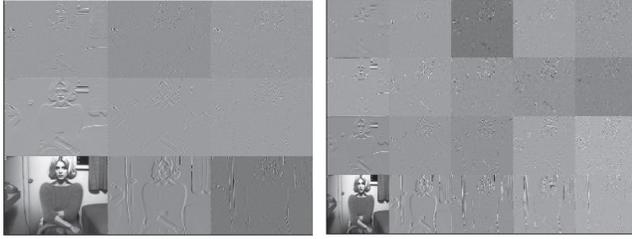
This technique provides a degree of flexibility which relates to the choice of resolution factors being potentially independent between different levels of decomposition. With respect to classic time-frequency representations, such as wavelets, polynomial basis decompositions do not necessarily use a dyadic partition and are therefore more adaptable.

Two examples of a first level decomposition on the same image are shown in Figure 1 with a decomposition using a  $3 \times 3$  Chebychev complete basis (left) and a  $5 \times 4$  Hermite complete basis(right).

### 4. BASE PROJECTIONS FOR FACIAL EXPRESSION RECOGNITION

The orthonormal polynomial decomposition allows to extract the different frequency components of a signal and offers the possibility to use multiresolution piecewise polynomial decomposition, so it can be used for the feature extraction within a system of facial expression recognition.

We use as input to our approach still face images labeled with landmarks around fiducial points. According to the difference of recording environment, the recorded data may



**Fig. 1.** Two examples of a first level decomposition with Chebychev 3x3 and Hermite 5x4 complete basis

contain different facial locations and scales. To eliminate such variation, we normalize each face. This is done by a global Procrustes analysis (GPA), followed by a histogram equalization.



**Fig. 2.** Example of input images with fiducial points

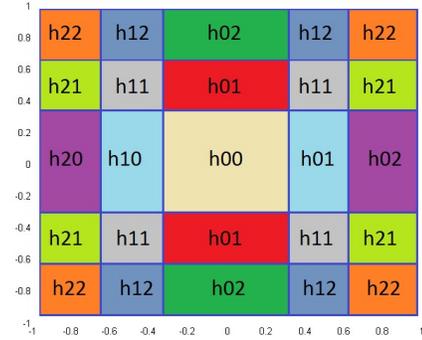
To extract the facial feature we propose to calculate the coefficients of polynomial projections on a complete basis on each fiducial point. Two different modes of computation are available : coefficients can either be calculated on texture patches, or retrieved from a multi-resolution polynomial decomposition.

For the first mode - **SR\_Poly**, feature vector for each facial point is extracted from a 19x19 pixels image patch centered on that point. This size was chosen to be similar to the size calculated empirically for the approach using LBP histograms. Hence, the polynomial coefficients are obtained via projections on a 19x19 complete Hermite basis with a Chebychev function for the collocation points. Since the coefficients provide a hierarchical representation of image structures, we can reduce their number to speed-up the computations with little efficiency loss.

For the second mode- **MR\_Poly**, we use a 3 level multi-resolution approach proposed in section 3. To have a similar representation to Gabor wavelets as [4] we use a complete 3x3 Hermite basis with a Chebychev function for collocation points. In this way, we will have a representation with 3 scales and 9 orientations. The regions around every fiducial point vary from 81x81 pixels to 3x3 pixels. Figure 3 shows the first level frequency decomposition of a 3x3 polynomial approach.

### Comparison with Gabor transform

Polynomial representations are similar to complete wavelet packet decompositions for a defined scale. Using a



**Fig. 3.** Frequency decomposition of the polynomial transform, where  $h_{i,j}$  represent different subbands

multi-resolution polynomial approach we can obtain a multi-scale/multi-orientation non-redundant representation.

Usually, in an automatic facial expression recognition system using Gabor wavelets, a bank of Gabor filters, composed of filters in distinct orientations and frequencies, is applied to the face to extract the feature vector. The filter bank is usually composed of four frequencies and six orientations. So to calculate the Gabor feature vector, each image is convolved with 24 Gabor kernels, which sizes varies with the frequencies. This representation is memory and time consuming. For example to calculate the 6 different orientations for the biggest Gabor kernel are required  $6 \times n \times n$  multiplications,  $n$  being the size of the kernel.

By using the polynomial projections with a 3x3 complete basis, our image patch is partitioned in 9 subpatches at each step, being considered as "orientations". Hence, the multi-scale polynomial transform will be more compact than a Gabor wavelet representation, thus allowing the disappearance of most sampling problems, such as the trade-off between orientation sampling and spatial sampling.

## 5. EXPERIMENTAL RESULTS

In our classification experiments we use a 10-fold cross-validation Support Vectors Machine-based approach on Cohn- Kanade and MUG databases.

The CohnKanade database [13] consists of expression sequences of 210 adults, starting from a neutral expression and ending in the peak of the facial expression. Participants were instructed by an experimenter to perform a series of 23 facial displays, six of which were prototypical emotions including angry, disgust, fear, joy, sad and surprise. We use a subset of 115 subjects for our experiments. Only the first (neutral) and final image (the prototypical expression) of each of the selected sequences are considered for our training and testing. The confusion matrix obtained from this database using multi-resolution polynomial projections is presented in Table 1. We see that happy, disgust, neutral, and surprise are detected with high accuracy while fear is presenting slightly inferior detection.

truth pred(%)	An	Di	Fe	Ha	Ne	Sa	Su
Anger	86.4	3.4	4.2	1.4	1.0	8.3	0.0
Disgust	2.3	93.2	4.2	1.4	1.0	0.0	0.0
Fear	2.3	1.7	79.2	2.9	0.0	4.2	1.2
Happiness	4.5	0.0	4.2	94.3	0.0	0.0	0.0
Neutral	0.0	1.7	0.0	0.0	92.9	4.2	1.2
Sadness	4.5	0.0	4.2	0.0	5.1	83.3	0.0
Surprise	0.0	0.0	4.2	0.0	0.0	0.0	97.6

**Table 1.** Confusion matrix for the CK database (MR\_Poly)

The MUG database [12] includes image sequences of 86 subjects performing the six basic expressions more than once. The image sequences begin and end at neutral state and follow the onset, apex, offset temporal pattern. For our experiments we used 401 images of 26 subjects that are manually annotated with 65 landmarks by removing the chin landmarks. The confusion matrix obtained from this database using multiresolution polynomial projections is presented in Table 2. In this case neutral presents the lowest detection, followed by fear. This is due to the low quantity of neutral images in the database. All others emotions were predicted with high accuracy.

truth pred(%)	An	Di	Fe	Ha	Ne	Sa	Su
Anger	94.6	1.4	0.0	0.0	3.6	4.0	0.0
Disgust	0.0	98.6	2.0	0.0	0.0	0.0	0.0
Fear	0.0	0.0	83.7	0.0	10.7	2.0	3.1
Happiness	0.0	0.0	2.0	100.0	10.7	0.0	1.5
Neutral	3.6	0.0	4.1	0.0	67.9	4.0	0.0
Sadness	1.8	0.0	0.0	0.0	7.1	88.0	1.5
Surprise	0.0	0.0	8.2	0.0	0.0	2.0	93.8

**Table 2.** Confusion matrix for the MUG database (MR\_Poly)

### Comparative Study

A comparison of the proposed methods with Gabor wavelets and LBP based texture descriptions [3] is shown in Tables 3 and 4. Table 3 shows the comparison results in terms of classification accuracy, and Table 4 in terms of execution time. All the experiments were carried on a Dell desktop with 2.53 GHz Intel Xeon CPU. The time given for feature extraction is for one single fiducial point.

Regarding the XY positions our results differ from the one presented in [4]. This is explained by the fact that our XY position are normalized by GPA, hence they are likely to give better results than Gabor wavelets.

Comparing our multiresolution polynomial approach to the one using Gabor wavelets, our method gives better performance results both in terms of accuracy as in terms of computation time. However, because the multiresolution polynomial approach implies the computation of coefficients which are unlikely to be relevant for classification, we will prefer a single-resolution method with coefficients pre-selections. As it turns out, multi-resolution decompositions are better for applications such as lossy compression or denoising than they are for classification.

Methods	Classification Rates(%)	
	Cohn Kanade	MUG
XY Positions	92.80	91.02
Gabor Wavelets	91.79	91.76
MR_Poly	92.03	92.27
LBP based method [3]	96.76	94.01
SR_Poly	94.54	93.25

**Table 3.** Comparison of proposed approaches with other methods in terms of classification accuracy

Methods	Execution times(ms)	
	Feature extraction	Classification
XY Positions	$\approx 0$	0.539
Gabor Wavelets	19.637	6.810
MR_Poly	5.575	7.303
LBP based method [3]	0.673	30.071
SR_Poly	0.411	1.420

**Table 4.** Comparison of proposed approaches with other methods in terms of execution times. It can be observed that even if SR\_poly performs with a slightly inferior precision ( $\approx 2\%$ ), it appears to be much better in terms of computation times w.r.t. LBP based method

It can be also observed that in comparison to the LBP based method while we obtain a slightly inferior precision ( $\approx 2\%$ ), our method appears to be much better in terms of computation times (over twenty times faster). The length of the feature vector extracted by LBP histograms is more substantial (59 uniform patterns for each fiducial point) so in terms of classification this method is time consuming.

## 6. DISCUSSION AND CONCLUSION

In this paper, we have proposed a new method of using coefficients obtained by polynomial projections for recognition of expressions from still face images.

We have shown that polynomial multi-resolution decomposition allows hierarchical organization of image information within the frequency domain. As a result, polynomial coefficients can be used as an efficient alternative to global or redundant texture representations such as Gabor Wavelets, without losing accuracy. Because polynomials in the complete basis are orthogonal, it is possible to compute the coefficients directly by a simple inner product of polynomials with the image. In multi-scale complete basis decompositions, while perfect reconstruction of the original signal can be obtained using a full set of coefficients, scalable approximation is also possible, by restricting reconstruction to a reduced set of coefficients, resulting in a fully scalable process.

Experimental results confirm that our approach performs well with face expression recognition, giving high accuracy results and being computationally efficient. In further works, we will study use of polynomial coefficients for the texture analysis within AAM models.

## 7. REFERENCES

- [1] T. Wu, S. Fu, and G. Yang, "Survey of the facial expression recognition research," in *Advances in Brain Inspired Cognitive Systems*, pp. 392–402. Springer, 2012.
- [2] V. Bettadapura, "Face expression recognition and analysis : the state of the art," *arXiv preprint arXiv :1203.6722*, 2012.
- [3] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns : A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.
- [4] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*. IEEE, 1998, pp. 454–459.
- [5] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression : machine learning and application to spontaneous behavior," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 2, pp. 568–573.
- [6] B.W. Char and S.M. Watt, "Representing and characterizing handwritten mathematical symbols through succinct functional approximation," in *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*. IEEE, 2007, vol. 2, pp. 1198–1202.
- [7] F. Ernawan, N.A. Abu, and N. Suryana, "Spectrum analysis of speech recognition via discrete tchebichef transform," in *2011 International Conference on Graphic and Image Processing*. International Society for Optics and Photonics, 2011, pp. 82856L–82856L.
- [8] DG Ciric and VD Pavlovic, "Linear phase two-dimensional fir digital filter functions generated by applying christoffel-darboux formula for orthonormal polynomials," *Electronics and Electrical Engineering*, vol. 120, no. 4, pp. 39–42, 2012.
- [9] I. Sadeh, "Polynomial approximation of images," *Computers & Mathematics with Applications*, vol. 32, no. 5, pp. 99–115, 1996.
- [10] M. Eden, M. Unser, and R. Leonardi, "Polynomial representation of pictures," *Signal Processing*, vol. 10, no. 4, pp. 385–393, 1986.
- [11] B. Augereau and P. Carré, "Hypercomplex polynomial wavelet packet application for color image," in *AGACSE 2012 Proceedings - Applied Geometric Algebras in Computer Science and Engineering (AGACSE 2012)*, La Rochelle : France, 2012.
- [12] N. Aifanti, C. Papachristou, and A. Delopoulos, "The mug facial expression database," in *Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 11th International Workshop on*. IEEE, 2010, pp. 1–4.
- [13] 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 *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.
- [14] AS Blaivas, "Visual analysis in unspecialized receptive fields as an orthogonal series expansion," *Neurophysiology*, vol. 6, no. 2, pp. 168–173, 1974.