A complete set of rotationally and translationally invariant features for images

Risi Kondor
risi@cs.columbia.edu
Computer Science Department, Columbia University,
1214 Amsterdam Ave., New York, NY10027, USA

Abstract

We propose a new set of rotationally and translationally invariant features for image or pattern recognition and classification. The new features are cubic polynomials in the pixel intensities and have the unusual property that up to numerical error and a bandwidth limit they are complete, in the sense that they uniquely determine the original image modulo rigid transformations. Our construction is based on the generalization of the concept of bispectrum to the three-dimensional rotation group SO(3), and a projection of the image onto the sphere.

Summary

The generalization of the classical bispectrum to a locally compact Lie group G takes the form

$$b(\rho_1, \rho_2) = C^\dagger \left(\widehat{f}(\rho_1) \otimes \widehat{f}(\rho_2) \right)^\dagger C \bigoplus_\rho \widehat{f}(\rho), \tag{1}$$

where ρ_1 and ρ_2 range over a complete set of inequivalent irreducible complex valued matrix representations of G, $\widehat{f}(\rho)$ are the corresponding Fourier components, and C is the Clebsch-Gordan matrix decomposing $\rho_1 \otimes \rho_2$ into a sum of irreducible representations here indexed by ρ. The bispectrum is of interest in signal processing because it is invariant to the action of G on f.

Kakarala proved the remarkable theorem that for compact Lie groups and some exceptional non-compact ones, as well as their homogeneous spaces, the bispectrum is complete, in the sense that it uniquely defines f, up to the action of G. Unfortunately, the group $\mathrm{ISO}^+(2)$ of rigid body motions in $\mathbb{R}^2$ is not compact, so we cannot directly apply this result to get translation and rotation invariant features for images.

Our paper proposes getting around this problem by projecting the image onto the sphere and taking advantage of the local isomorphism between the action of $\mathrm{ISO}^+(2)$ on a small region around the origin and the action of the rotation group SO(3) (which is compact) on a small region around the North pole of the sphere.

In the end we get a convenient set of invariants of the form

$$p_{l_1,l_2,l} = \sum_{m=-l}^{l} \sum_{m_1=-l_1}^{l_1} C^{l_1,l_2,l}_{m_1,m-m_1,m} \, \widehat{f}^{*}_{l_1,m_1} \, \widehat{f}^{*}_{l_2,m-m_1} \, \widehat{f}_{l,m},$$

where the $\widehat{f}_{l,m}$ are just spherical harmonic coefficients, and the $C^{l_1,l_2,l}_{m_1,m-m_1,m}$ are "ordinary" Clebsch-Gordan coeficients, which are easy to compute. If u is the size of the original image in pixels, the space requirement of the bispectrum is $O(u^{3/2})$, while the time complexity of computing it is $O(u^{5/2})$.

We tried out the bispectral features on a dataset of randomly rotated and translated MNIST digits. While the baseline method of RBF kernels is stumped by these random transformations, an SVM trained on bispectral features performs consistently well. For more details and references see `http://arxiv.org/abs/cs.CV/0701127`.