Fast, Illumination Insensitive Face Detection Based on Multilinear Techniques and Curvature Features

Christian Bauckhage
Deutsche Telekom Laboratories
10587 Berlin, Germany
christian.bauckhage@telekom.de

Thomas Käster
Bielefeld University
33615 Bielefeld, Germany
tkaester@techfak.uni-bielefeld.de

Abstract

This paper brings together two recent developments in image analysis. We consider a new mathematical framework that provides illumination invariant descriptors for face detection. Towards fast learning and processing, we understand images and the corresponding feature maps as multilinear entities and apply higher order classifiers for image analysis and object detection. Experimental results underline that this approach indeed provides quick training, fast runtime and robust performance across a variety of illumination conditions.

1. Introduction

Varying scene illumination and ambient lighting still very much affect the performance of most present day computer vision systems. Recently, Koenderink [4] claimed that this is due to methodical flaws in mathematical image modeling and proposed a representation providing illumination invariance. More recently still, Bauckhage and Tsotsos [1] applied this framework in face detection. Concerned with feature vectors according to Koenderink's ideas, they found that even simple, linear subspace techniques can cope with considerable illumination variations. Nevertheless, the results in [1] are of little practical use.

The major shortcoming is that the authors focus on PCA-based classification. For finding faces in an image, however, this is rather inefficient. With d denoting the dimension of the subspace used for classification, principal component analysis of all possible subimages of size $m \times n$ requires $O(dmn)$ operations per pixel. This simple analysis underlines, that, for common image resolutions and most choices of m and n, naïve linear techniques may be suited for *recognition* but not for *detection*.

In this paper, we propose a much faster approach to illumination insensitive detection. Similar to the work in [1], we consider curvature features computed according to Koenderink's theory. However, instead of linear techniques we apply multilinear classifiers.

Treating image patches as *higher order tensors* or n*-way arrays* leads to interesting results [5, 6, 7, 9, 10]. In short, the findings reported in these recent contributions suggest that multilinear techniques capture salient structures more efficiently and more faithfully than conventional linear approaches. We investigate if this also applies to illumination insensitive face detection. First, we briefly sketch Koenderink's framework and a novel approach to multilinear classification. Then, in practical experiments, we combine both approaches and obtain results showing that multilinear face detection based on curvature features performs fast and robustly. A conclusion will end this paper.

2. Image Space $\mathbb{I}^3$ and Higher Order Classifiers

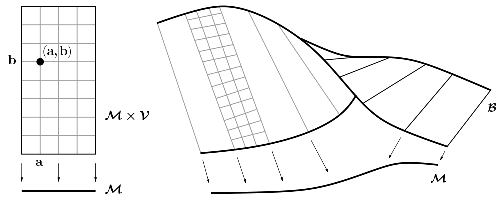
In this section, we summarize Koenderink's approach to illumination invariant image processing and briefly introduce multivariate representations for object detection.

2.1. Image Space $\mathbb{I}^3$

In [4], Koenderink criticizes that grayscale images often are taken to be entities embedded in $\mathbb{R}^3$. If intensity values z_i were the surface of some function over the image plane, i.e. $z_i = f(x_i, y_i)$, the geometry of $\mathbb{R}^3$ would allow for arbitrary rotations of this surface. However, some such rotations might cause intensity values to lie in the coordinate plane and image coordinates to become parallel to the intensity direction.

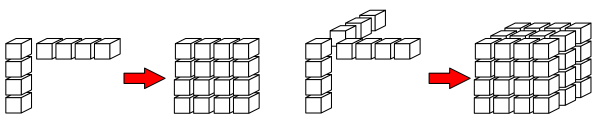
Seeking a model that prevents physically senseless configurations, Koenderink proposes the use of fibre bundles (see Fig. 1). His image space $\mathbb{I}^3$ locally looks like $\mathbb{P}^2 \times \mathbb{L}$ where the base manifold $\mathbb{P}^2$ corresponds to the picture plane and the fibers $\mathbb{L}$ represent intensity information. Moreover, arguing that the photon count on a CCD chip is Poisson distributed, Koenderink stipulates a log-intensity scale

0-7695-2521-0/06/$20.00 (c) 2006 IEEE



(a) (b)

Figure 1. A fibre bundle $\mathcal{B}$ is defined by a mapping $\pi : \mathcal{B} \rightarrow \mathcal{M}$. $\mathcal{M}$ is the base manifold and $\mathcal{V}_x = \pi^{-1}(x)$ are the fibers; $\mathcal{B} = \bigcup_{x \in \mathcal{M}} \mathcal{V}_x$. Locally, $\mathcal{B}$ resembles $\mathcal{M} \times \mathcal{V}$.



(a) $\vec{u} \otimes \vec{v}$ (b) $\vec{u} \otimes \vec{v} \otimes \vec{w}$

Figure 2. The outer product of two or several vectors results in higher order objects.

$Z(x, y) = \log(z(x, y)/z_0)$ where z_0 is an arbitrary unit of intensity. Images thus correspond to cross sections of $\mathbb{I}^3$ and an image point is a triple $\{x, y, Z\}$.

Since, in $\mathbb{I}^3$, intensity direction and image plane cannot mix, brightness transformations do not alter relations among fibres. This leaves the curvature of cross sections invariant. Finally, due to the bundle structure, Gaussian- and Mean curvature of cross sections are given by notably simple expressions. In contrast to the lengthy formulas known from Euclidean geometry, they simply correspond to

$$K(x, y) = \frac{\partial^2 Z}{\partial x^2}\frac{\partial^2 Z}{\partial y^2} - \frac{\partial^2 Z}{\partial x \partial y} = Z_{xx} Z_{yy} - Z_{xy}^2 \qquad (1)$$

and

$$H(x, y) = \frac{\frac{\partial^2 Z}{\partial x^2} + \frac{\partial^2 Z}{\partial y^2}}{2} = \frac{Z_{xx} + Z_{yy}}{2}. \qquad (2)$$

2.2. Classification of Multilinear Objects

As a digital image consists of one or several layers, for classification, it may be interpreted as a third-order tensor $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{m_1 \times m_2 \times m_3}$ where m_1 and m_2 correspond to the x- and y-resolution and m_3 counts the number of layers (usually $m_3 \leq 3$). Alas, detection or recognition algorithms, which make use of linear algebra, usually treat image patches $\boldsymbol{\mathcal{X}}$ of size $m \times n \times d$ as high dimensional vectors $\vec{x} \in \mathbb{R}^{mnd}$. However, for classifiers which are based on an inner product, one can refrain from unfolding $\boldsymbol{\mathcal{X}}$ by considering the inner product of tensors $\boldsymbol{\mathcal{W}} \cdot \boldsymbol{\mathcal{X}} = \sum_{ijk} \mathcal{W}_{ijk} \mathcal{X}_{ijk}$.

(a) grayscale (b) gradient (c) H and K curvature in $\mathbb{I}^3$

Figure 3. Examples of training data for second and third order multilinear classification.

Dealing with a two class problem such as face detection, multilinear classification closely resembles the linear case. With ω_+ denoting the class of face images and ω_- denoting the class of all non-face images, the decision function is given by:

$$\omega(\boldsymbol{\mathcal{X}}) = \begin{cases} \omega_+ & \text{if} \quad \boldsymbol{\mathcal{W}} \cdot \boldsymbol{\mathcal{X}} > \theta \\ \omega_- & \text{otherwise} \end{cases}$$

Moreover, if the projection tensor $\boldsymbol{\mathcal{W}}$ is given as a sum over R tensors of rank-1, such that

$$\boldsymbol{\mathcal{W}} = \sum_{r=1}^{R} \boldsymbol{\mathcal{W}}^r = \sum_{r=1}^{R} \vec{u}^r \otimes \vec{v}^r \otimes \vec{w}^r \qquad (3)$$

where $\otimes$ denotes the outer product (see Fig. 2) and $\vec{u}^r \in \mathbb{R}^m$, $\vec{v}^r \in \mathbb{R}^n$, $\vec{w}^r \in \mathbb{R}^d$, image analysis can be done efficiently. Applying the classifier $\boldsymbol{\mathcal{W}}$ to an image $\boldsymbol{\mathcal{I}}$ reduces to a sequence of one-dimensional convolutions

$$\boldsymbol{\mathcal{I}} * \boldsymbol{\mathcal{W}} = \sum_{r=1}^{R} \boldsymbol{\mathcal{I}} * \boldsymbol{\mathcal{W}}^r = \sum_{r=1}^{R} \big((\boldsymbol{\mathcal{I}} * \vec{u}^r) * \vec{v}^r \big) * \vec{w}^r. \qquad (4)$$

Therefore, the effort is $O(R(m + n + d)) \ll O(dmn)$, if $R \ll \min\{m, n\}$ and $m, n \gg d$.

If a sample $\{\boldsymbol{\mathcal{X}}^l, y^l\}_{l=1,\dots,L}$ of image patches $\boldsymbol{\mathcal{X}}^l$ and corresponding class labels y^l is given, $\boldsymbol{\mathcal{W}}$ can be found by minimizing the least squares error

$$\begin{aligned} E &= \sum_{l=1}^{L} \Big(y^l - \boldsymbol{\mathcal{X}}^l \cdot \sum_{r=1}^{R} \vec{u}^r \otimes \vec{v}^r \otimes \vec{w}^r \Big)^2 \\ &= \sum_{l=1}^{L} \Big(y^l - \sum_{r=1}^{R} \sum_{ijk} \mathcal{X}_{ijk}^l u_i^r v_j^r w_k^r \Big)^2 \end{aligned} \qquad (5)$$

As there is no closed form solution for (5), an iterative procedure is required. Typically, a gradient descend with

$$\frac{\partial E}{\partial u_i^r} = -2 \sum_{l} \Big(y^l - \sum_{r} \sum_{ijk} \mathcal{X}_{ijk}^l u_i^r v_j^r w_k^r \Big) \Big(\sum_{jk} \mathcal{X}_{ijk}^l v_j^r w_k^r \Big)$$

and corresponding $\frac{\partial E}{\partial v_j^r}$ and $\frac{\partial E}{\partial w_k^r}$ will find a locally optimal solution. However, if the problem is cast as a sequence of convex optimizations, there is a more efficient way of finding a suitable projection tensor. Consider the following alternating least squares procedure:

1. initialize $\vec{u}(0) \in \mathbb{R}^m$ and $\vec{v}(0) \in \mathbb{R}^n$, $\|\vec{u}\| = \|\vec{v}\| = 1$
2. given $\vec{u}(t)$ and $\vec{v}(t)$, solve the least squares problem
$$\vec{w}(t) = \underset{\vec{w}}{\operatorname{argmin}} \sum_l \big(y^l - \vec{w} \cdot \sum_{ij} \mathcal{X}^l_{ijk} u_i v_j\big)^2$$
$$= \underset{\vec{w}}{\operatorname{argmin}} \sum_l \big(y^l - \vec{w} \cdot \vec{x}^l_k\big)^2$$
for conventional least squares problems like this, there is a closed form solution: $\vec{w}(t) = \big(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\big)^{-1}\boldsymbol{X}^{\mathrm{T}}\vec{y}$ where $\boldsymbol{X} = [\vec{x}^1_k, \ldots, \vec{x}^L_k]^{\mathrm{T}}$ and $\vec{y} = [y^1, \ldots, y^L]^{\mathrm{T}}$
3. correspondingly, given $\vec{u}(t)$ and $\vec{w}(t)$, solve for $\vec{v}(t)$; normalize $\vec{v}(t)$ to unit length
4. correspondingly, given $\vec{v}(t)$ and $\vec{w}(t)$, solve for $\vec{u}(t)$; normalize $\vec{u}(t)$ to unit length
5. while $\|\vec{u}(t) - \vec{u}(t-1)\| > \tau$, continue with 2.

Compared to gradient-based approaches, this algorithm trains faster, for it requires less tensor-tensor and tensor-vector multiplications and optimizes in much lower dimensional spaces. It works, because, in each iteration, it reduces the overall error E. Moreover, the sequence $\{\vec{u}(t)\}_{t\in\mathbb{N}}$ lies on the unit ball in $\mathbb{R}^m$ which is a compact convex set. The sequence must therefore have a convergent subsequence and the algorithm is guaranteed to find a local minimum. If the tensor $\boldsymbol{\mathcal{W}}$ is constrained to be orthogonally decomposable [11], extending the procedure to an R-term solution is straightforward. After finding a set of vectors $\vec{u}^r$, $\vec{v}^r$ and $\vec{w}^r$, the vectors for the next term $\vec{u}^{r+1}$, $\vec{v}^{r+1}$ and $\vec{w}^{r+1}$ are required to be orthogonal to the ones found so far.

3. Experimental Results

This section explores multilinear classifiers and curvature features in $\mathbb{I}^3$ for fast face learning and rapid face detection under varying ambient illumination. Our investigation comprises multilinear representations of different orders, different features as well as an analysis of the influence of the number of rank-1 tensors for classifier design.

All experiments were conducted on a 1.8GHz Pentium Mobile Notebook running LINUX. The evaluation set consists of 310 gray level images (scaled to 320×240 pixels) of the Yale face database [3]. This subset corresponds to the subsets 1, 2 and 3 proposed by Georghiades et. al. [3]. It covers 10 individuals under 31 different illuminations. For training, 4 images were randomly selected for each illumination condition. Afterwards, 5 positive and 20 negative example patches of size 100×100 pixels were extracted from

method	t_{train}	t_{test}	EER
Gray (R = 2)	5s	13s	55%
Gray (R = 4)	13s	12s	83%
Gray (R = 6)	18s	12s	87%
Gradient (R = 2)	10s	16s	80%
Gradient (R = 4)	18s	16s	80%
Gradient (R = 8)	31s	16s	88%
$\mathbb{I}^3$ Features (R = 2)	18s	24s	88%
$\mathbb{I}^3$ Features (R = 4)	37s	23s	90%

Table 1. Quantitative results.

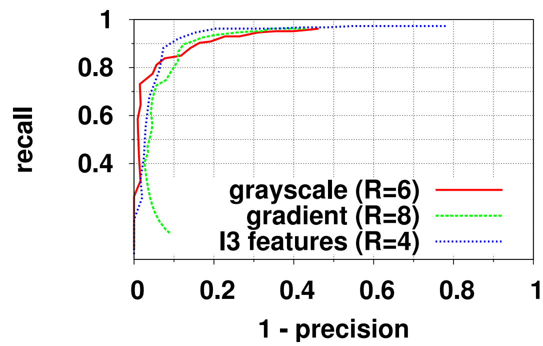


Figure 4. Precision/recall curves of the best classifiers for each of the different features.

each of the 124 training images. Each multilinear discriminant classifier was thus trained with $\omega_+ = 620$ positive and $\omega_- = 2480$ negative image patches.

In order to better asses the use of Koenderink's features, we also experimented with simple gray value and common gradient magnitude images (in both cases, for classification, face images were treated as second order tensors in $\mathbb{R}^{100\times100}$). In the $\mathbb{I}^3$ based experiments, Gaussian- and Mean curvature maps K and H were combined into third order tensors in $\mathbb{R}^{100\times100\times2}$. Figure 3 shows examples of all three representations. For computing gradient and curvature maps, we applied a fast and precise operator introduced by Deriche [2].

Table 1 shows how multilinear classification performs for the different feature types. It lists training- and runtimes as well as equal error rates (EERs) for projection tensors of different numbers R of terms. The EERs result from precision/recall curves (see Fig. 4) which were obtained by varying the classification threshold θ. It is noticeable that, for the different features, the best performances (highlighted in grey) were obtained for different Rs. In terms of EER peak performance, the third order classifier applied to combined Mean- and Gaussian-curvature achieves best results.

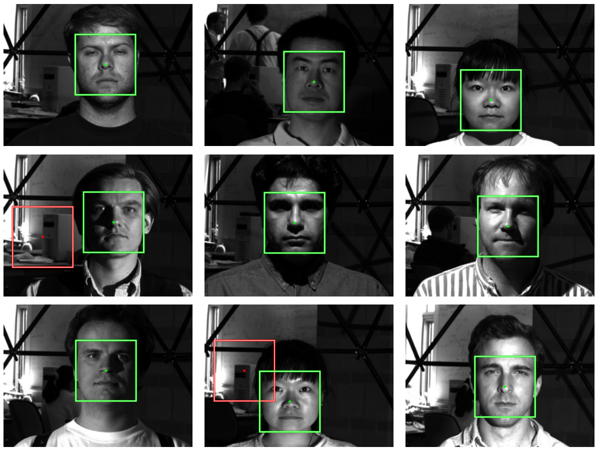


Figure 5. Exemplary detection results for the third order tensor classifier using $\mathbb{I}^3$ features.

Adding more rank-1 tensors to the corresponding $\mathcal{W}$ did not lead to further improvements.

Concerning the average execution speed, Tab. 1 underlines the quick learning and fast runtime capabilities of separable multilinear classifiers. Here, the alternating least squares approach discussed in section 2 displays its potential: even the computationally most demanding classifier requires less than a minute for training. The increasing runtimes in the third column of the table reflect the additional effort due to computing partial derivatives. Nevertheless, even the classifier based on H and K curvature maps reaches a processing speed of 8Hz. Figure 5 exemplifies detection results this classifiers yields for different, rather extreme configurations of light sources.

For baseline comparison, we also considered the state of the art algorithm by Viola and Jones [8]. Trained on the same training set and applied to the same test set, this boosted predictor achieved an EER of 92%. However, training took 93 minutes – two orders of magnitude longer than for the third order multilinear classifier. Therefore, while EER performance and runtimes are comparable, in terms of training effort multilinear classifiers applied to curvature tensors outperform boosted weak classifiers. With regard to illumination insensitive face detection, $\mathbb{I}^3$ features and multilinear classification thus open up interesting perspectives for scenarios where adaptivity is an asset.

4. Conclusion

Recent research has demonstrated that tensor-based classifiers robustly capture essential image structures. Moreover, as they are separable, rank-1 decomposable tensor classifiers can be trained rapidly and allow for fast processing of image data. In this paper we explored their use for face detection under diverse lighting conditions. To this end, we considered illumination insensitive curvature feature maps resulting from Koenderink's approach to image modeling [4]. Similar to the results reported in [1], we found that curvature features computed in image space $\mathbb{I}^3$ enable robust face detection across a wide range of different ambient lighting. However, in contrast to that contribution, the tensor-based classifiers explored in this paper trained within seconds and provided runtimes of several Hz.

In conclusion, applying multilinear classifiers to $\mathbb{I}^3$ features yield fast and robust performance where face detection has to cope with changing and inhomogeneous illumination. It thus provides an auspicious approach for a wide range of practical applications. Encouraged by the fast training times, we currently explore applying our approach to scenarios where online learning may overcome problems due to uncontrollable and constantly changing scene illumination. In particular, we are interested in advanced interaction with mobile devices such as cell phones.

References

[1] C. Bauckhage and J. Tsotsos. Image Space I3 and Eigen Curvature for Illumination Insensitive Face Detection. In *Proc. Int. Conf. on Image Analysis and Recognition*, volume 3656 of *LNCS*, pages 456–463. Springer-Verlag, 2005.

[2] R. Deriche. Recursively Implementing the Gaussian and Its Derivatives. In *Proc. ICIP*, pages 263–267, 1992.

[3] A. Georghiades, P. Belhumeur, and D. Kriegman. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE Trans. Pattern Anal. Machine Intelli.*, 23(6):643–660, 2001.

[4] J. J. Koenderink and A. J. van Doorn. Image Processing Done Right. In *Proc. ECCV*, volume 2350 of *LNCS*, pages 158–172. Springer, 2002.

[5] A. Shashua and T. Hazan. Non-Negative Tensor Factorization with Applications to Statistics and Computer Vision. In *Proc. Int Conf. Machine Learning*, pages 792–799, 2005.

[6] A. Shashua and A. Levin. Linear Image Cosing for Regression and Classification using the Tensor-rank Principle. In *Proc. CVPR*, volume I, pages 42–40, 2001.

[7] M. Vasilescu and D. Terzopolos. Multilinear Analsysis of Image Ensembles: Tensorfaces. In *Proc. ECCV*, volume 2350 of *LNCS*, pages 447–460. Springer, 2002.

[8] P. Viola and M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. In *Proc. CVPR*, volume I, pages 511–518, 2001.

[9] H. Wang and N. Ahuja. Compact representation of multidimensional data using tensor rank-one decomposition. In *Proc. ICPR*, volume I, pages 44–47, 2004.

[10] S. Yan, D. Xu, L. Zhang, X. Tang, and H.-J. Zhang. Discriminant Analysis with Tensor Representation. In *Proc. CVPR*, volume I, pages 526–532, 2005.

[11] T. Zhang and G. Golub. Rank-One Approximation To High Order Tensors. *SIAM J. Matrix Anal. Appl.*, 23(2):534–550, 2001.