To appear in: Proc. ICASSP-08, Las Vegas, NV, Feb 2008.

CONTEXTUALLY ADAPTIVE SIGNAL REPRESENTATION USING CONDITIONAL PRINCIPAL COMPONENT ANALYSIS

Rosa M. Figueras i Ventura[1], *Umesh Rajashekar*[1], *Zhou Wang*[2], *Eero P. Simoncelli*[1]

[1] HHMI and New York University, USA [2] University of Waterloo, Canada

ABSTRACT

The conventional method of generating a basis that is optimally adapted (in MSE) for representation of an ensemble of signals is Principal Component Analysis (PCA). A more ambitious modern goal is the construction of bases that are adapted to *individual* signal instances. Here we develop a new framework for instance-adaptive signal representation by exploiting the fact that many real-world signals exhibit local self-similarity. Specifically, we decompose the signal into multiscale subbands, and then represent local blocks of each subband using basis functions that are linearly derived from the surrounding context. The linear mappings that generate these basis functions are learned sequentially, with each one optimized to account for as much variance as possible in the local blocks. We apply this methodology to learning a coarse-to-fine representation of images within a multi-scale basis, demonstrating that the adaptive basis can account for significantly more variance than a PCA basis of the same dimensionality.

Index Terms— Adaptive basis, conditional PCA, self-similarities, image modeling, image representation.

1. INTRODUCTION

A fundamental concept in signal processing is that of selecting a representation that is optimally adapted to a class of signals. In the most well-known formulation of this problem, known as Principal Components Analysis (PCA), a linear basis is selected such that each successive axis captures as much variance (estimated over the signal ensemble) as possible [1]. Although it is optimal over the entire signal class, this basis may perform poorly on individual examples.

Recent literature explores the problem of adapting a basis to individual signal instances. Specifically, one attempts to select a subset of basis elements from a large redundant dictionary of functions such that the signal is best represented. Although the general problem is NP-hard, it can be solved in some special cases, and a variety of algorithms have been developed to find approximate solutions, [*e.g.,* 2, 3, 4] that have been used for both compression, [*e.g.,* 5, 6] and denoising, [*e.g.,* 7, 8]. Nevertheless, these methods tend to be computationally expensive, and compression examples to date have not demonstrated consistent or significant advances over traditional methods, primarily due to the cost of encoding the indices of selected basis elements.

An interesting property found in many naturally-occurring signals is self-similarity: local structures occur repeatedly at different locations, orientations and scales within a given signal instance. Neither PCA nor sparse adaptive bases exploit this property, but a number of authors have developed alternative methods of doing so. Fractal image coding schemes encode each imaged block as a transformed copy of a larger block, where transformations include translation, rotation, shearing, and contraction [9]. State-of-the-art texture synthesis methods use a related strategy, generating novel texture images by carefully stitching together blocks randomly selected from within an example texture image, [*e.g.,* 10, 11]. Image self-similarities have also been exploited for image repairing/inpainting, [*e.g.,* 12], and for image denoising, [*e.g.,* 13, 14, 15].

In this paper, we present a method of learning an adaptive basis that is generated from the signal itself. Specifically, each patch of signal is represented using basis functions that are linearly computed from contextual information. Patches from one band of a subband image decomposition use as conditional neighborhood patches on the same position in a coarser subband. A related representation has been previously developed for image distortion analysis [16]. Note that this approach differs from the fractal coding approach, because the contextual information is drawn from a *fixed* neighborhood around each patch. In addition, we do not attempt to directly represent the patch contents, but instead to generate a *basis* for the patch. We develop a method of optimizing the matrix that generates the basis functions, so as to maximize the variance accounted for in patches of an ensemble of signals. As such, we refer to this method as *conditional PCA* (cPCA). We apply cPCA to the problem of image representation, exploiting the self-similarity that occurs between co-located patches in adjacent scales of a multi-scale decomposition. We show that this representation is capable of capturing significantly more variance than PCA over the same patches, and examine the degradation that occurs when this conditional representation is cascaded from coarse scales to fine scales.

2. CONDITIONED PCA

2.1. Sequential PCA learning problem

The Principal Component Analysis (PCA) problem corresponds to learning a sequence of basis functions that best capture the variance of an ensemble of signals [1]. Given N signals, represented as vectors $\{\mathbf{y}_n | n = 1, 2, \cdots, N\}$ of dimension K, the first PCA basis vector represents an axis that captures the one-dimensional signal subspace with highest variance, the second captures the second highest variance axis, and so on and so forth. Although the PCA basis may be directly computed as the eigenvectors of the signal covariance matrix, ordered according to the eigenvalues, it may also be computed sequentially by optimizing one component at a time:

$$\mathbf{b}_k = \arg\min_{\mathbf{b}} \sum_{n=1}^{N} \min_{s_{k,n}} \|\mathbf{y}_n^{(k)} - s_{k,n}\mathbf{b}\|^2 \qquad (1)$$

where we define

$$\begin{aligned} \mathbf{y}_n^{(1)} &= \mathbf{y}_n, \\ \mathbf{y}_n^{(k)} &= \mathbf{y}_n^{(k-1)} - s_{k-1,n}\mathbf{b}_{k-1}, \qquad k > 1, \\ &= \mathbf{y}_n - \sum_{j=1}^{k-1} s_{j,n}\mathbf{b}_j. \end{aligned}$$

Due to the orthogonality of PCA basis functions, each optimal coefficient $s_{k,n}$ needs to be computed once oh the k^{th} iteration. This greedy formulation produces exactly the same result (up to numerical considerations) as the eigenvector solution.

2.2. Sequential cPCA learning problem

We base our formulation of the cPCA problem on the PCA formulation described above, but instead of solving for a fixed set of basis vectors, we seek the best transformation of a set of input data vectors to a form a basis for a set of output data vectors. Consider N input-output vector data pairs $\{\mathbf{x}_n | n = 1, 2, \cdots, N\}$ and $\{\mathbf{y}_n | n = 1, 2, \cdots, N\}$, where the dimensions of $\mathbf{x}_n$'s and $\mathbf{y}_n$'s are M_1 and M_2, respectively. The objective is to find a set of operators $\{\mathbf{B}_k : R^{M_1} \rightarrow R^{M_2} | k = 1, 2, \cdots, K\}$ such that the K-dimensional subspaces spanned by vectors $\{\mathbf{B}_k(\mathbf{x}_n)\}$ captures as much of the variance of the vectors $\{\mathbf{y}_n\}$ as possible. In order to reduce the complexity of the problem, we assume the operators are linear (i.e., the $\mathbf{B}_k$'s are $M_2 \times M_1$ matrices). Even with this restriction, fixed nonlinear transformations may be incorporated by applying them to the contextual data and appending these results onto the original input vectors x_n.

Following the greedy PCA formulation of the previous section, we seek to minimize:

$$\mathbf{B}_k = \arg\min_{\mathbf{B}} \sum_{n=1}^{N} \min_{s_{k,n}} \left\| \mathbf{y}_n^{(k)} - s_{k,n}\mathbf{B}\mathbf{x}_n \right\|^2 . \qquad (2)$$

Unlike the PCA formulation, the basis vectors $\{\mathbf{B}_j\mathbf{x}_n | j = 1 \ldots k\}$ are not guaranteed to be orthogonal, and thus, we must compute each $\mathbf{y}_n^{(k)}$ by eliminating (projecting out) the subspace spanned by all previously generated basis vectors $\{\mathbf{B}_j\mathbf{x}_n | j = 1 \ldots k-1\}$ from the original $\mathbf{y}_n$:

$$\begin{aligned} \mathbf{y}_n^{(k)} &= \mathbf{y}_n - \sum_{j=1}^{k-1} s_{j,n}^{(k)} \mathbf{b}_{jn}, \\ \{s_{j,n}^{(k)}\} &= \arg\min_{\{s_{j,n}\}} \sum_{n=1}^{N} \sum_{j=1}^{k-1} \|\mathbf{y}_n - s_{j,n}\mathbf{B}_j\mathbf{x}_n\|^2 . \qquad (3) \end{aligned}$$

The $\mathbf{B}_k$'s are computed sequentially (greedily), starting with $k = 1$. For each k, the $\{s_{j,n}^{(k)}\}$ are computed from Eq. (3) by pseudo-inverting the current adapted basis set $\{\mathbf{B}_j\mathbf{x}_n | j = 1 \ldots k-1\}$ for each n (this set is empty when $k = 1$). Then the $\mathbf{B}_k$'s are obtained from Eq. (2), by alternating between solving for the optimal $\mathbf{B}_k$ and solving for the optimal $s_{k,n}$. Although we cannot prove convexity of the overall objective function, this alternation is a form of coordinate descent and thus guaranteed to converge to a (possibly local) minimum.

3. APPLYING CPCA TO IMAGES

In this section, we show preliminary tests illustrating the potential advantages of the cPCA representation for image coding. Image structures exhibit several forms of self-similarity, with respect to translation and dilation. Thus, we will apply cPCA within a multiscale, oriented transform in which these self-similarities are easily recognized and represented. Specifically, we use a steerable pyramid decomposition [17], which partitions the image into oriented subbands of spatial frequency varying by factors of two. This representation is an overcomplete tight frame, and we use it because the contents of each scale are translation-invariant and rotation-invariant.

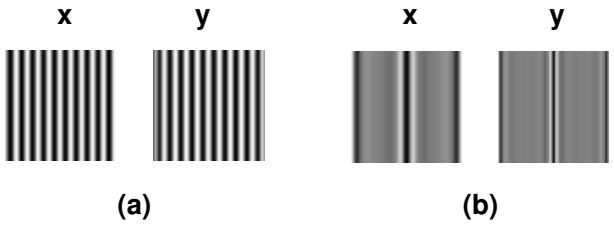


Fig. 1. Self-similarity between co-located non subsampled coarse-scale (x) and fine-scale (y) patches, for (a) an oscillating texture and (b) a step edge.

For our examples, we use a steerable pyramid with three orientation bands, which is roughly five-fold overcomplete. Both the cPCA and the PCA basis used in this section are computed for fixed size patches of a given subband.

Selecting our vectors $\mathbf{y}_n$ to be spatial blocks of coefficients within a subband, we wish to choose context vectors $\mathbf{x}_n$ based on the content of co-located blocks at the next coarsest ("parent") subband at the same orientation. In studying the relationship between these child-parent blocks, we find that different image structures exhibit different behaviors. Specifically, edges produce oscillations at roughly the center frequency of each subband, but textures produce oscillations that are locked to the signal frequency in both subbands. This behavior is illustrated in Fig. 1. This means that textures and periodic signal structures will appear with the same frequency in the (non-downsampled) parent as in the child, but edges will appear at a frequency in the child that is twice that of the (non-downsampled) parent. Thus, we define the conditioning vectors, $\mathbf{x}_n$, to include the contents of a (non-downsampled) block from the parent band, concatenated with those of a frequency-doubled version of that block.

Figure 2 illustrates four typical examples of reconstructing 11x11 patches from a horizontal "child" band by two methods: first, using four PCA basis functions of the child band, and second, using four cPCA basis functions that are computed from the parent band. Note that the PCA basis functions for all patches are fixed (by design), but the basis functions of cPCA, which are computed by applying fixed transformations $\{\mathbf{B}_k | k = 1, \ldots 4\}$ to the parent vector $\mathbf{x}_n$, are different for each child patch. Loosely speaking, we find that the cPCA basis functions appear to be similar to one of the two parent patches, or slightly translated copies of those patches.

Figure 2(a) shows an example where the child patch has a strong horizontal structure. Since the first four PCA basis functions capture predominantly horizontal structures, the patch reconstructed using the PCA basis functions is a satisfactory representation of the child patch. Similar results are obtained using the cPCA basis. Figure 2(b) shows a case where the PCA reconstruction captures the horizontal structure in the child, but fails to represent the non-horizontal ones. cPCA, on the other hand, is capable of capturing this signal change. Figure 2(c) illustrates a striking example in which the child patch is oriented obliquely (the subband filters in the steerable pyramid are broadly tuned for orientation). This is poorly represented by the PCA basis, but the cPCA basis exploits the structure in the parent to provide a high-quality reconstruction of the child patch. Finally, in Fig. 2(d), we show that even when the child patch does not have any predominant orientation, cPCA is able to reconstruct the child patch more faithfully than PCA.

As we have seen, wavelet bands do tend to retain structure from coarse to fine, but in some cases, structures that cannot be seen in the coarser band appear in finer bands. In such a case, it is impossible to recover anything from linear combinations of the parent. We can

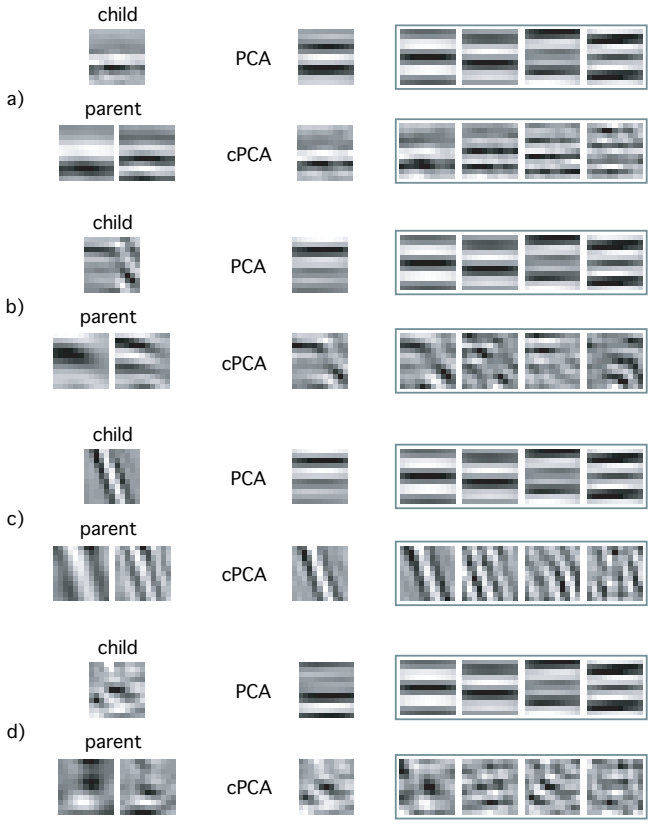


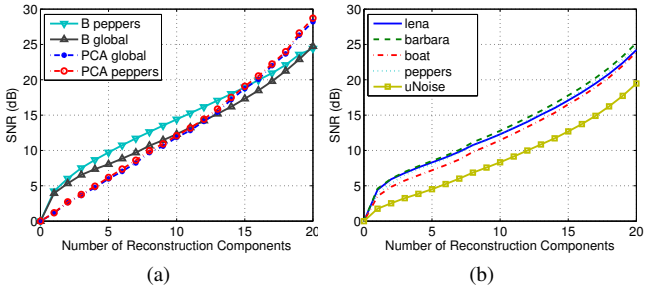
Fig. 2. Representation of four example patches using PCA and cPCA bases. The first column shows the subband patches. The "child" is the patch $\mathbf{y}_n$ to be approximated, and the "parents" are a patch from the same location of the (non-downsampled) coarser frequency band and its frequency-doubled version, the concatenation of which forms $\mathbf{x}_n$ in the cPCA equations. The second column (single patch) shows approximations of different patches obtained with 4-component PCA and cPCA bases ($\mathbf{y}_n^{(4)}$). Finally, the third column (groups of four patches) shows the four different basis functions used for the PCA reconstruction on the top ($\{\mathbf{b}_k,\ k = 1 \ldots 4\}$, basis functions globally derived from the child band, identical for all three examples) and the four basis functions used for the cPCA reconstruction ($\{\mathbf{B}_k\mathbf{x}_n,\ k = 1 \ldots 4\}$ basis functions adaptively derived from the concatenated parents, obtained with the same four linear prediction matrices for all rows).

alleviate this problem with a simple modification of Eq. (2), adding a fixed basis function, $\mathbf{b}_k$, to the conditionally-adapted basis function:

$$\{\mathbf{B}_k, \mathbf{b}_k\} = \arg\min_{\mathbf{B},\mathbf{b}} \sum_{n=1}^{N} \min_{s_{k,n}} \left\| \mathbf{y}_n^{(k)} - s_{k,n} \left(\mathbf{B}\mathbf{x}_n + \mathbf{b}_k\right) \right\|^2 . \quad (4)$$

$\mathbf{b}_k$ allows the basis to capture possible "new arrivals" in the child band that cannot be predicted from the parent. Eq. 4 is equivalent to PCA when $\mathbf{B} = 0$, and to Eq. (2) when $\mathbf{b} = 0$. Notice that the PCA and cPCA basis components share a common coefficient, and thus this formulation does not incur any additional representational cost.

The cPCA matrices $\mathbf{B}_k$ can be optimized over the image being represented (as in the example of Fig. 2), or over a separate training set. Of course, even when they are derived from a generic training set, they still produce different basis functions for every patch. The



(a) (b)

Fig. 3. SNR versus number of components for standard PCA and cPCA with a standard PCA component (block size 5×5, pyramid with three orientations and three scales). (a) Performance in encoding one subband of the "peppers" image, with different bases, including: cPCA with $\mathbf{B}_k$s optimized on "peppers" (downward triangles); cPCA optimized on a separate training set of three images (upward triangles); PCA optimized on "peppers" (dashed); and PCA optimized on the separate training set (circles), almost overlapping the previous one. (b) Performance in encoding a subband of different images with bases optimized over a separate training set. Notice that all images have approximately the same behavior, except for the white noise, which, as expected, is more difficult to approximate.

latter is preferable for coding applications, so that the $\mathbf{B}_k$'s need not be encoded in the data stream. Figure 3(a) shows the SNR behavior of cPCA when the linear transform has been trained on a set of three images (Lena, Barbara and Boat) on 5×5 patches, and applied to an image not in the training set (peppers). This figure also shows the performance of cPCA when the linear transform $\mathbf{B}$ is trained on the image being represented (peppers), and when the linear transform has been learned from white noise (which captures the redundancy of the steerable pyramid basis only). Also shown are the results of PCA, both trained on the peppers image, and trained on the separate training set. The cPCA representation outperforms PCA for small numbers of basis elements by roughly 4dB. Once roughly half of the components of the basis have been used for the reconstruction, cPCA loses its advantage with respect to PCA. Nevertheless, the remaining error at this stage is quite small. The error drops to zero when all 25 basis functions are included (not shown in plot).

Figure 3(b) shows only the cPCA behavior for a set of several images, including white noise. The matrices $\mathbf{B}_k$ are optimized over the same three-image training set mentioned previously. The behavior of cPCA is quite consistent over all the images, and does not vary much for images in or out of the training set. The only curve that falls significantly short of the general behavior is that of the noise image, which does not exhibit the same self-similarity properties as the photographic images. We have also examined these same curves for images represented with orthogonal wavelet transforms, to determine whether the learned $\mathbf{B}$ was capturing redundancies due to overcompleteness in the steerable pyramid, or redundancies due to image self-similarity. We find that the behavior of cPCA relative to PCA on orthogonal wavelets is similar (although absolute performance is worse), demonstrating that the primary advantages we report here are due to image self-similarity.

Finally, Fig. 4 shows an example of an image reconstructed from only a single PCA or cPCA basis function in every 5x5 patch of the two highest-frequency pyramid levels, along with the high-frequency (non-oriented) residual band. Both the PCA basis and the cPCA matrices have been trained on a set that did not contain the image under consideration, thus there is no associated cost for encoding either

(a) PCA, PSNR=25.99dB (b) cPCA, PSNR=27.26dB

Fig. 4. Reconstruction of an image by substituting the last two steerable pyramid levels and the residual high frequency band with (a) a one-term PCA approximation, and (b) a one-term cPCA cascaded representation. See text.

basis. In order to do the PCA and cPCA approximations, the pyramid bands are divided in non-overlapping 5x5 patches, and every patch is represented with a single basis function. For cPCA, this is done recursively from coarse to fine scale, computing each band (except for those at the coarsest scale) from the previously approximated parent (and thus, accumulating the errors). PCA is performed independently for each subband. Despite the cascading of errors, the one-component cPCA approximation behaves better than the one-component PCA approximation, both visually and in PSNR.

4. CONCLUSIONS

We have introduced conditional PCA, a new methodology for adaptive signal representation that directly exploits contextual self-similarities. This representation offers potential advantages for signals in which the variance of a local patch (or coefficient block) conditioned on its context is significantly less than the unconditioned variance.

Specifically, we believe cPCA offers several advantages over recently developed methods for adaptive basis selection from finite redundant dictionaries: 1) the cPCA basis is not limited to a finite set, and thus has the potential to adapt to a much larger variety of local image structures (assuming they are self-similar); 2) The main computational cost is the learning of the $\mathbf{B}_k$'s, which can be done offline from a training set. Computation of the locally adaptive bases is then linear, and thus far more efficient than current approximate methods for selecting an optimal basis from a redundant dictionary; and 3) the basis functions are not specified directly, but are computed from the signal itself. For example, in compression, the signal specific basis functions could be computed from previously received data, and in denoising, they can be computed from previously cleaned data.

We have shown preliminary examples of image representation, which demonstrate the potential usefulness of cPCA. In this context, we are currently working to improve the algorithm by which the $\mathbf{B}_k$'s are optimized, and we are exploring alternative choices of conditioning neighborhoods, including for example surrounding coefficients in the same band, or coefficients from other orientation bands. More generally, we believe the cPCA representation can offer advantages in many other applications where PCA or any other signal-adapted image representation is used. This includes a broad range of computer vision, pattern recognition and image processing applications, such as compression, restoration, super-resolution, and object recognition.

5. REFERENCES

[1] I. T. Jolliffe, *Principal Component Analysis*, Springer series in Statistics. Springer, 2 edition, 2002.

[2] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. on Signal Proc.*, vol. 41, no. 12, pp. 3397–3415, 1993.

[3] N. G. Kingsbury and T. H. Reeves, "Redundant representation with complex wavelets: how to achieve sparsity," in *ICIP*, 2003, vol. 1, pp. 45–48.

[4] Michal Aharon, Michael Elad, and Alfred Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. on Signal Proc.*, vol. 54, pp. 4311–4322, 2006.

[5] E. Le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *IEEE Trans. on Image Proc.*, vol. 14, no. 4, pp. 423–438, 2005.

[6] R. M. Figueras i Ventura, P. Vandergheynst, and P. Frossard, "Low-rate and flexible image coding with redundant representations," *IEEE Trans. on Image Proc.*, vol. 3, pp. 726–739, 2006.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[8] J.-L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. on Image Proc.*, vol. 11, no. 6, pp. 670–684, 2002.

[9] M. F. Barnsley and L. P. Hurd, *Fractal image compression*, A. K. Peters, Ltd, 1993.

[10] J. De Bonet and P. Viola, "A non-parametric multi-scale statistical model for natural images," in *Adv. in Neural Info. Processing*. December 1997, vol. 9, MIT Press.

[11] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. of the Int. Conf. on Computer Vision*, Sept. 1999, vol. 2, p. 1033.

[12] Z. Wang, Y. L. Yu, and D. Zhang, "Best neighborhood matching - an information loss restoration technique for block based image coding systems," *IEEE Trans. on Image Proc.*, vol. 7, pp. 1056–1061, 1998.

[13] Z. Wang and D. Zhang, "Restoration of impulse noise corrupted images using long-range correlation," *IEEE Signal Processing Letters*, vol. 5, pp. 4–6, 1998.

[14] A. Buades, B. Coll, and J. M. Morel, "Image denoising by non-local averaging," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 2005, vol. 2, pp. 25–28.

[15] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. on Image Proc.*, vol. 16, no. 16, pp. 2080–2095, 2007.

[16] Z. Wang and E. P. Simoncelli, "An adaptive linear system framework for image distortion analysis," in *Proc. of ICIP.*, 2005, vol. 3, pp. 1160–1163.

[17] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation," in *ICIP*, 1995, vol. 3, pp. 444–447.