The Steerable Pyramid: Wavelets in Vision and Image Processing

Imdad Rizvi#, Krishnamoorthy V Iyer## and Sandip Kothari##

Abstract—This is a review paper on the steerable pyramid, a linear multiscale, multiorientation image decomposition that finds use in the preprocessing stage of image processing and computer vision applications. Orthogonal separable wavelet transforms, once popular for the same tasks, give rise to heavily aliased representations that do not represent diagonal orientations well. To overcome these limitations, the steerable pyramid was proposed. The steerable pyramid is partly inspired by visual information processing in area V1 of mammalian cortex of the brain, which uses a strategy wherein individual cells respond maximally to a small range of orientations and receptive field (RF) sizes, but the cellular population as a whole has cells responding to all orientations and RF sizes.

Index Terms—Aliasing, directional derivative, edge detection, early vision, steerable pyramid, orientation, orthogonal separable wavelet.

I. INTRODUCTION

There has long been a fruitful exchange of ideas between neuroscientists working on information processing in the brain, and engineers, computer scientists, and mathematicians working in fields such as machine vision/computer vision, who seek to understand and/or replicate the phenomenal success of biological systems in processing information about their environment, on the other.

This influence runs both ways – the field of artificial neural networks was loosely based on the knowledge of neural circuitry then current. In the other direction, the application of information theory (to take just one example) to neuroscience has already yielded promising new results, such as enabling us to formulate new hypotheses regarding neural information processing.

The attempt to subject these hypotheses to the test of experiment has given rise to several new experimental paradigms. To take another example, it has been suggested that the sensory systems of most vertebrates and probably a number of invertebrates apply wavelet-lie transforms, with the mammalian visual system being the focus of the most intense studies [1]. The aforementioned is an obvious case where the influence of mathematics and engineering can be clearly seen in the field of neuroscience. In this paper, we will look at a case where the influence runs in the reverse direction, from neuroscience to engineering, to wit, computer vision and image processing.

This paper discusses another kind of nonseparable two dimensional linear image transform, the steerable pyramid, which is inspired partly by current knowledge of visual cortex. There also exist applications of the steerable pyramid to image processing and computer vision, which lie beyond the scope of this paper, but which we will mention in passing.
This is a review paper and organized as follows: Section II, based on [3], introduces very briefly the different kinds of linear image transforms and the criteria to be used in choosing a linear transformation for coding purposes. Section III, based on [4], discusses the use of the 1D Haar transform to analyzing a 2D data set, viz. an image, two different ways of applying the Haar to an image and the limitations of each. Section IV, based on [5], introduces the solution suggested – shiftable multiscale transforms. The problem of aliasing is not peculiar only to the orientation domain, the same issue arises when the signal is translated or dilated. The concept of “joint shiftability” is introduced. In Section V, the concept of “shiftability”, specialized to the orientation domain, and known as “steerability”, is considered, based on [6]. (Note that the concept of “shiftability” was introduced as a generalization of the concept of “steerability” and the sampling theorem). In Section VI, we discuss, based on [7], [8], the steerable pyramid, and also mention in passing possible applications, though we do not have space to consider these in detail. Section VII concludes the paper.

II. SUBBAND TRANSFORMS

Simoncelli and Adelson [3] consider the two Broad kinds of Linear transforms used in image processing, image analysis, and image coding. These are of two kinds: transform coders and subband coders. Though the distinction is not clearcut, the authors describe transform coding as those techniques which are based on orthogonal linear transforms, the classic example being the Discrete Fourier Transform (DFT), where a signal is decomposed into sinusoidal frequency components. On the other hand, subband transforms arise by convolving the input signal with a set of bandpass filters and decimating the results. Each subband encodes a specific portion of the frequency spectrum (and this corresponds to information at a particular spatial scale). However, there exists significant overlap between transform coders and subband coders: the block Discrete Cosine Transform (DCT) is an example. The authors then go on to consider the criteria that must be taken into account in choosing a linear transformation for coding purposes. In their view, basis filters must show spatial frequency localization (achieved by bandpass filters tuned to a particular scale and orientation – we will see that the steerable pyramid achieves precisely this simultaneous localization of scale and orientation) and spatial localization – in other words, the transform should encode positional information. This requirement for joint localization in the spatial and spatial frequency domains leads us quite naturally to consider the use of wavelet or wavelet like transforms. The final property the authors consider is orthogonality between the basis functions. The Haar transform, discussed in the next section, is an example of an orthogonal transform. A consequence of imposing the condition of orthogonality is that they have the same number of transform coefficients as pixels, as a consequence of which the image is said to be maximally decimated or critically sampled [2]. But the condition of orthogonality (which is attractive from a computational perspective, as it minimizes the number of basis functions required to represent the function and still retain complete information) leads us to other issues such as aliasing and noninvariance of the transform coefficients to translations, dilations, and rotations of the input signal, issues which will be addressed in section IV, where we discuss the shiftable multiscale transform that was developed to deal with these difficulties.
III. ORTHOGONAL PYRAMID TRANSFORMS

Adelson and Simoncelli [4] discuss the application of what may be the simplest orthogonal pyramid transform to image coding, to wit, the Haar wavelet. We shall discuss this in some detail, as it provides a tutorial introduction to other, more complicated 2D transforms, such as nonseparable transforms like the steerable pyramid. The authors were motivated by the desire to find a linear image transform whose basis functions are localized in both space and spatial frequency, and also satisfy the property of orthogonality. (As mentioned in the introduction to their paper, and as emphasized elsewhere in the paper as well, this is motivated by evidence that the human visual system performs a similar image decomposition in its early processing and to emulate some of its properties). In pursuit of this, they chose the Haar wavelet transform. It is true that the Haar wavelet transform has very good localization in space but very poor localization in (spatial) frequency. (This limitation is addressed by the authors’ recent work, which we discuss below in later sections). Specifically, the paper discusses the construction of a Haar basis set for an (toy) image of eight pixels. They go on to discuss two different ways of applying the Haar transform. In the first method, a 2D Haar transform is computed by combining a pair of 1D Haar transforms “separably”. By that we mean that the image is considered as being composed of a column, each element of which is a row of horizontal data, to each of which the 1D Haar transform is applied. The resultant is then considered to be a row, each element of which is considered to be a column of vertical data, and again the Haar transform is applied to this. Unfortunately, this procedure leads to basis functions of widely varying shapes, which violate the principle of self similarity (the requirement that the basis functions should all have essentially the same shape but should be dilated versions of each other).

A more attractive (in the sense that it violates fewer of the requirements for an image transform, as discussed at the end of the previous section) 2D basis set is also discussed, in which, instead of separably computing the Haar transform in the vertical and horizontal directions separately, we form a separable Haar pyramid by recursively subdividing the 2D low frequency information at each step. In this case, there are only 3 shapes of kernels, and they are repeated at various positions and scales. In separable pyramid, at each stage of the recursion, the lowpass information is divided into four further channels: lowpass, horizontal, vertical, and “diagonal”. The vertical and horizontal channels are tuned for orientation as well as spatial frequency. As discussed in [4], the separable application of the 1D Haar transform to images results in a representation in which one of the subbands, the “diagonal” subband contains a mixture of two orientations. (Note that this “diagonal” channel is tuned for both orientations of diagonal high frequency, and so is different from the diagonal orientations to which the steerable pyramid – to be discussed below – is sensitive via its bandpass basis functions). This paper also introduces some of the other 2D quadrature mirror filter (QMF) pyramids such as the quincunx and hexagonal pyramids (also discussed in [2]), but we will not discuss these in this paper, as our goal is to discuss the steerable pyramid. (In passing, we would like to speculate on the possibility that the compound eyes of many insects, which are arranged in a hexagonal pattern, may make use of the properties of the hexagonal pyramid in the analysis of their visual environments – of course only experimental work can validate/invalidate such hypotheses). A fundamental issue with all orthogonal transforms is that the
orthogonality of the transform leads to critical sampling, which in turn leaves these representations noninvariant to shifts of the input signal – this problem arises in the translation, scale and orientation domains. This necessitates the introduction of shiftable multiscale transforms [5], discussed in the next section.

IV. SHIFTABLE MULTISCALE TRANSFORMS

This section considers the most important shortcoming of orthogonal wavelet transforms and the resulting critically sampled representations: their noninvariance to “shifts” (see below) of the input signal. To overcome this shortcoming, the authors of [5] demonstrate the need for nonorthogonal wavelet transforms and using the resulting overcomplete representations. The primary motivation for wavelet transforms as a class is to obtain simultaneous localization in the time and frequency domains. When applied to 2D, wavelet transforms may be computed using basis functions that are localized for space (i.e. position) and spatial frequency (which, in the log polar domain, corresponds to dilation and rotation of the image in the spatial domain).

Now we turn to the main point of this section: the lack of translation invariance of orthogonal wavelet transforms (in the spatial domain), with similar lack of invariance to changes in scale and orientation (in the spatial frequency domain) and how this necessitates defining the twin pair of concepts known as “shiftability” and “joint shiftability.” The basis functions of a wavelet transform are related by translations and dilations. Translations and dilations correspond to two of the most common “physical” transformations of a signal. Due to the relationship between basis functions, it might be expected that the coefficients of the transform would behave in a “simple” manner when the input is translated or dilated.

But this does not happen. The authors go on to discuss why that does not happen and illustrate their point by referring to the example of spatial translation and applying a discrete dyadic Daubechies wavelet kernel of length four. The example illustrates that translation invariance, which is of course a desirable property for a transform (any transformation) to have, is not attained for this transform. Translation invariance cannot be attained in a system which uses convolution and subsampling. In other words, translation of the input signal cannot produce a corresponding simple translation of the transform coefficients, unless the translation is a multiple of each of the subsampling factors in the system. A weaker form of translation invariance is obtained if all the information in a subband remains within the subband as the input signal is translated. This happens if the Nyquist criterion is satisfied. But a critically sampled transform (such as orthogonal wavelet transforms are) violates the Nyquist criterion, and hence the information moves from one subband to another, thus violating even the simple form of translational invariance above. The authors prove this in their paper. In the process, they formalize the notion of “shiftability” and “joint shiftability” which respectively refer to the invariance under shifts in the input to one or more parameters.

The three primary shiftability notions introduced for a 2D transform (for which the basis functions are position, orientation, and scale) are shiftability in position, shiftability in orientation, and shiftability in scale. Although each of these parameters are continuous, we do not want an infinite set of basis functions, each tuned for slightly different parameter values. The issues of
shiftability arise because we want to employ only a finite basis set and yet interpolate responses corresponding to any point in parameter space from the coefficients corresponding to a finite set of samples. They then discuss the mathematical conditions that are necessary and sufficient for shiftability to hold. In the process of the discussion with regards to orientation, the authors also refer to a paper by Freeman and Adelson [6], where the concept of shiftability in the orientation domain ("steerability") was first introduced. (See the next section). As with the translation domain, we wish to be able to interpolate measurements at all orientations by means of a finite set of filters. The authors also introduce the idea of steerability of the Fourier transforms of the basis functions. They explain the significance by considering a polar separable frequency domain function as the product of a function H of polar angle and a function U of radial distance. Shiftability of the function H corresponds to "steerability" of the frequency domain function. One can also define steerability for functions that are not polar separable. It is with the application of the shiftability concept to scale that certain difficulties arise, which the authors address in a short section. In introducing the concept of "joint shiftability", the authors note that for independent parameters, joint shiftability can easily be achieved. But for parameters that are Fourier complements of each other, difficulties arise. Scale and position are Fourier complements of each other. The authors prove that a one dimensional subsampled transform with basis functions parametrized for position and scale cannot achieve shiftability.

The authors discuss the application to image processing tasks such as stereo matching and image enhancement. Note that stereo matching is precisely the kind of problem that primates and carnivores (such as lions and tigers) must solve in order to survive. (In the case of carnivores, for example, they must form an accurate estimate of the "depth" of prey in their visual field in order to make an accurate spring). We will briefly summarize the authors' discussion of stereo matching. The visual system is confronted with two views of a scene from differing positions (left and right eyes). The stereo matching task consists of finding the relative horizontal displacement between corresponding points in two images – a displacement that is inversely proportional to the distance to the point in the three dimensional world. Performing this matching enables the animal to form a "depth map" from the two images. When the displacements of corresponding points between the two images is small, the matching task is simplest. As a result of this, stereo matching algorithms are applied within a multiresolution representation. The images are first matched at coarse spatial scales, where all displacements are small relative to the distance between pixels, and the results are then used as initial estimates for increasingly smaller scales. In the next section, we discuss very briefly the paper that introduced the concept of "steerability" (shiftability in the orientation domain).

V. STEERABLE FILTERS

In [6], Freeman and Adelson introduced the concept of steerability for oriented filters. The significance of their paper lies in their analysis of the role of oriented filters. "Steerable filter", as used by the authors, refers to a class of filters in which a filter of arbitrary orientation is synthesized as a linear combination of a set of "basis filters". The importance of this lies in the fact that this is a more efficient approach to calculating the response of a filter at a different orientation rather than having a large number of filters, one for each
orientation. What does it mean to say that a function of two spatial variables is “steerable”? It means that the function can be written as a linear combination of rotated copies of itself. The authors discuss and derive the conditions necessary to “steer” a given filter and derive various theorems. They emphasize that all functions that are bandlimited in angular frequency are steerable, given enough basis filters. But the most useful filters are those that require a small number of basis filters. The authors go on to discuss designing steerable filters. Two possibilities may arise: the basis functions are not all xy separable, which can increase the computational cost. For machine vision applications, one would like to have xy separable basis functions. They note that the steerable filter transform as described by them leads to overcomplete representations. Such representations incur increased computational requirements but bring other advantages, such as those discussed in the previous section. The authors end with a discussion of applications of the use of steerable filters, among which are its uses for analyzing local orientation, edge detection, and shape from shading.

VI. THE STEERABLE PYRAMID

We now discuss two papers [7], [8], each of which deals with the actual design of a steerable pyramid filter. The authors describe various constraints the filters must satisfy. There are three mathematical constraints to be satisfied in order that perfect reconstruction be possible and one angular constraint in order that the obtained filter be steerable. These the authors discuss, and in addition they discuss the recursive procedure that they develop and employ to design the filters to satisfy the constraints. To summarize, a set of filters forms a steerable basis if they are rotated copies of each other and also if a copy of the filter at any orientation may be computed as a linear combination of the basis filters. The scale tuning of the filters is one of the three conditions that must be satisfied for perfect reconstruction to be possible. This transform has steerable orientation subbands, and is “selfinverting” (i.e., the matrix corresponding to the inverse transformation is equal to the transpose of the forward transformation matrix) and is aliasing free. The pyramid can be designed to produce any number of orientation bands. If the number of orientation bands is $k$, the resulting transform will be overcomplete by a factor of $4k/3$. They also go on to compare the properties of the steerable pyramid transform with other multiscale decompositions such as the Laplacian pyramid and the separable orthogonal wavelet transforms discussed in the previous sections.

VII. CONCLUSION

Steerable pyramid is a linear multiscale, multiorientation image decomposition. It is widely used in the preprocessing stage of image processing and computer vision applications. Steerable Pyramid is also used in satellite remotely sensed imagery for image image fusion. The steerable pyramid overcome problem of heavily aliased representations.

REFERENCES


