MULTISCALE OBJECT FEATURES FROM CLUSTERED COMPLEX WAVELET COEFFICIENTS

Ryan Anderson, Nick Kingsbury, Julien Fauqueur

Signal Processing Group, Dept. of Engineering
University of Cambridge, UK
http://www-sigproc.eng.cam.ac.uk

ABSTRACT

This paper introduces a method by which intuitive feature entities can be created from ILP (InterLevel Product) coefficients. The ILP transform is a pyramid of decimated complex-valued coefficients at multiple scales, derived from dual-tree complex wavelets, whose phases indicate the presence of different feature types (edges and ridges). We use an Expectation-Maximization algorithm to cluster large ILP coefficients that are spatially adjacent and similar in phase. We then demonstrate the relationship that these clusters possess with respect to observable image content, and conclude with a look at potential applications of these clusters, such as rotation- and scale-invariant object recognition.

1. INTRODUCTION

Multiscale representations of images possess many advantages in object recognition and image retrieval activities. If an object has a known and simple multiscale profile - that is, a sparse set of feature entities in a known spatial and scale pattern - then the search and identification of transformed instances of a desired object is simplified. In particular, if one can first search a decimated search domain for a coarse level representation of an object, the potential exists to accelerate an object recognition algorithm.

The Dual-Tree Complex Wavelet (DT CWT) [1] has the ability to decompose a 2-D image into a decimated, multiscale representation that isolates coarse image components into a sparse set of equally spaced complex coefficients. In prior work [2], we have demonstrated a method by which this set of coefficients may be manipulated into a new set, the *Interlevel Product (ILP)*. The phases of the ILP consistently represent the type of feature in the spatio-scalar vicinity, where a feature type may be a step edge or a ridge. Thus, the presence of such a feature will result in a tight spatial cluster of large similar-phase ILP coefficients at the appropriate scale; and, given an observation of such coefficients, one may infer the presence and characteristics of the original feature.

In this paper, we seek to perform this latter task; from a set of coarse-scale ILP coefficients, we will infer the presence of abstract ILP "feature" entities. Each entity c will require the following parameters to represent feature characteristics:

- **Feature Type: θ_c.** A feature is either a pure ridge, a pure edge, or a combination of the two ("combination" is a loosely defined term that includes curvy features, half-ridges, noisy edges, etc.) The feature type is represented by the mean complex angle, θ_c, of the ILP coefficients that comprise the entity.
- **Feature Location: μ_c.** The 2-D location of the feature is defined relative to other features in its level of the multiscale hierarchy. Feature locations may also be defined relative to specific parent features in the next coarsest scale. And, ultimately, there will be a "root" feature against which all features are relatively measured. This feature will typically be the largest or most salient feature at the coarsest level.
- **Feature Shape: Σ_c.** The shape parameter summarizes the size, orientation, and spatial distribution of the entity. We adopt the parameter name Σ_c because, in this paper, our shapes are covariance matrices of Gaussian distributions. However, shape parameters may be much more flexible to accommodate correspondingly more complex shapes.
- **Feature Saliency: α_c.** The saliency of the feature refers to the level of contrast of the feature, and corresponds to the magnitude of the comprising ILP coefficients.

To create entities with these parameters, we will modify the traditional EM-trained Gaussian Mixture Model to cluster large, same-phase coefficients. The resultant entities

This work has been carried out with the support of the UK Data & Information Fusion Defence Technology Centre.

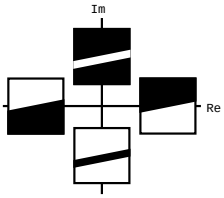


Fig. 1. Relationship between the complex phase of an ILP coefficient in the 15° subband and the nature of a $\sim 15^\circ$ feature in the vicinity. Note that this phase-to-feature relationship is constant in all subbands (with appropriately oriented features).

should not only be robust to multiscale misalignment (discussed further in [2]), but they should also possess semantic meaning with regard to visually identifiable image features.

We begin in section 2 with a more detailed description of the ILP transform with examples; section 3 describes the modified GMM routine we use, and section 4 shows the results. We conclude in section 5 with further discussion on the motivation for this methodology and future work.

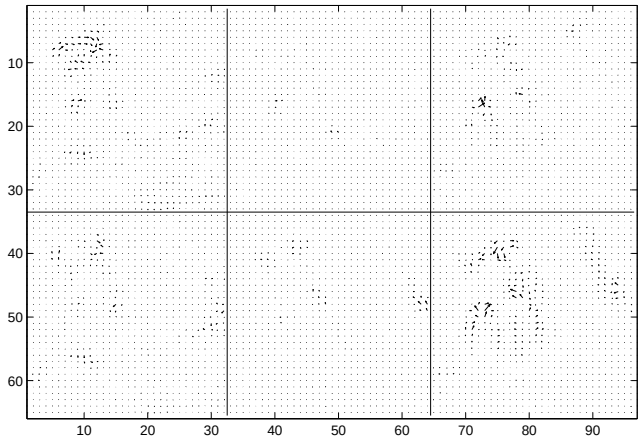
2. THE ILP TRANSFORM

For an $N \times M$ pixel image, the ILP is a pyramid of L levels, where each level $l = 1 \ldots L$ possesses $\frac{N}{2^l} \times \frac{M}{2^l} \times 6$ complex coefficients, where the six values at each spatial location correspond to six directional subbands at 15°, 45°, 75°, 105°, 135°, and 165°; thus, it is identical in dimension to the DT CWT upon which it is based. It is created by multiplying the DT-CWT coefficients at the corresponding location and level l by the complex conjugate of a phase-doubled, interpolated version of their $l + 1$ parent coefficients; this process is outlined in detail in [2].

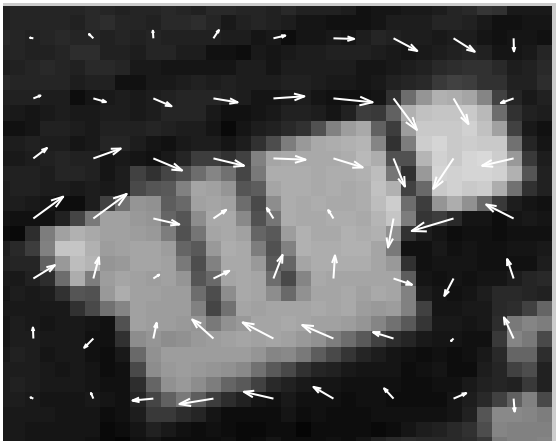
The phases of DT CWT coefficients are dependent upon the spatial offset of directional features, and are therefore unreliable to represent objects consistently. By contrast, the phases of the ILP are only influenced by the nature of the feature in the vicinity. Specifically, the relationship of phase with local feature type is shown in Figure 1; large-magnitude values that are purely real or purely imaginary indicate edges and ridges (respectively) that are high-contrast and ideal. Large magnitude ILP coefficients that possess both imaginary and real components indicate hybrid features (which can include curves, wide ridges, or ridge-edge combinations), and small magnitude coefficients indicate areas that are generally flat (where the phase is random and meaningless).

In Figure 2, one can see the six-subband level 2 ILP decomposition (b) of a typical aerial image (a). By applying the phase relationships shown in Figure 1, one can visually associate a macroscopic feature in the image with a cluster of corresponding ILP coefficients in the ILP representation.

(a)



(b)



(c)

Fig. 2. An aerial image (a) and its Level 2 ILP Decomposition, χ_2, shown in (b) with 15°, 45°, 75°, 105°, 135°, and 165° subbands respectively displayed clockwise from the top left corner. In (c), we zoom in upon the 15° subband coefficients, and overlay them upon the original image. Note how the top and bottom edges of the building are represented by complex coefficients with phases predicted by Figure 1.

For instance, in (c), one can consider the top edge of the upper-left white building as an abstract feature that corresponds to the cluster of ILP coefficients that are positive-real (i.e. pointing to the right) in the corresponding location of the 15° subband. These positive-real coefficients extend not only along the length of the edge, but are somewhat broad in the direction normal to the edge as well. This "softening" of edges and ridges in the ILP domain allows us to use smooth Gaussian clusters to represent these originally crisp features, provided that the cluster can be associated with a complex phase (which would be positive-real, in this case).

3. EM ALGORITHM FOR DIRECTIONAL CLUSTERING

The Gaussian Mixture Model builds a set of k Gaussian clusters upon i data points x, where each cluster $c = 1 \ldots k$ is weighted by a value, α_c, and possesses the following distribution:

$$p(x_i|\mu_c, \Sigma_c) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_c|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)} \quad (1)$$

The parameters μ, Σ, and α are determined through the iterative Expectation-Maximization algorithm [3] by which the log likelihood expression

$$\sum_{c=1}^{k}\sum_{i=1}^{N} \log(p(x_i|\mu_c, \Sigma_c))p(c|x_i, \Theta_c) \quad (2)$$

is maximized with respect to each parameter, while all other parameters remain fixed. We calculate $p(c|x_i, \Theta_c)$ using Bayes' Rule and α_c as a prior, and Θ_c refers to the complete parameter set $\{\mu_c, \Sigma_c, \alpha_c, \theta_c\}$ for all $c \in k$. The update parameters for the α_c cluster weights are

$$\alpha_c^{new} = \frac{1}{N}\sum_{i=1}^{N} p(c|x_i, \Theta_c) \quad (3)$$

Equation (1) is used to model single-instance data points at the given x_i locations. If we were to imagine that the data point x_i was independently and repeatedly observed in the same location β_i times, the joint probability of this event can be expressed as

$$p(x_i, \beta_i|\mu_c, \Sigma_c) = \frac{1}{Q_i}\left(p(x_i|\mu_c, \Sigma_c)\right)^{\beta_i} \quad (4)$$

where Q_i is a normalizing constant. If we allow β_i to take on any positive real value, rather than just integer multiples, we can also accommodate any data set where individual x_i locations are weighted by any value of β_i.

Our ILP data, however, is a set of continuous complex ILP coefficients from level l, χ_l, where x_i is the location of the $\chi_l^{(i)th}$ coefficient, and we need to create a real scalar value $\beta_{i,c}$ with respect to cluster c to use Equation (4). Our clusters must be designed such that the largest adjacent sets of these coefficients (in the same directional subband) with similar phase should be clustered together and represented by their mean phase, θ_c. Thus, we define $\beta_{i,c}$ to be the *projection* of the $\chi_l^{(i)th}$ coefficient onto a unit vector with the proposed cluster phase θ_c:

$$\beta_{i,c} = \Re\left[(\chi_l^{(i)})^* \cdot e^{j\theta_c}\right] \quad (5)$$

and $j = \sqrt{-1}$. We can now model these real-valued projections as $\beta_{i,c}$ instances of a single-instance data point at that location. Note, however, that these values will be negative if the difference between $\angle\chi_l^{(i)}$ and θ_c is greater than 90°. These negative values cause difficulty with the convergence of the EM algorithm, and are not perceptually sensible; we do not necessarily want a clustering algorithm to over-penalize a region that contains coefficients with opposing phase. Thus, we set a lower bound of zero upon $\beta_{i,c}$, which is equivalent to setting the magnitude of opposite-phase coefficients to zero.

Our expression for $p(x_i|\mu_c, \Sigma_c)$ has now become

$$p(x_i, \beta_{i,c}|\mu_c, \Sigma_c, \theta_c) = \frac{1}{Q_i}\left[\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_c|^{\frac{1}{2}}} e^{-\frac{1}{2}((x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)}\right]^{\beta_{i,c}} \quad (6)$$

Inserting these new values into equation (2) produces the following expression:

$$\sum_{c=1}^{k}\sum_{i=1}^{N}\left[\log\frac{1}{Q_i} - \frac{\beta_{i,c}}{2}\left[\log|\Sigma_c| + (x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)\right]p(c|x_i, \Theta_c)\right] \quad (7)$$

We maximize equation (7) with respect to Σ_c and μ_c by differentiating and setting the result to zero, and solving for each of these variables. The procedure for determining Σ_c and μ_c is almost identical to the original GMM derivation, which is explained comprehensively in [4]. The resulting iterative update equations for Σ_c and μ_c are shown below:

$$\mu_c^{new} = \frac{\sum_{i=1}^{N}\beta_{i,c}p(c|x_i, \Theta_c)x_i}{\sum_{i=1}^{N}\beta_{i,c}p(c|x_i, \Theta_c)}$$

$$\Sigma_c^{new} = \frac{\sum_{i=1}^{N}\beta_{i,c}p(c|x_i, \Theta_c)(x_i-\mu_c^{new})(x_i-\mu_c^{new})^T}{\sum_{i=1}^{N}\beta_{i,c}p(c|x_i, \Theta_c)}$$

From our interpretation of $\beta_{i,c}$ as multiple instances of a data point at x_i, we can see that the update equation for α^{new}, equation (3), may be rewritten as

$$\alpha_c^{new} = \frac{\sum_{i=1}^{N} \beta_{i,c} p(c|x_i)}{\sum_{i=1}^{N} \beta_{i,c}} \quad (8)$$

Note that we have abbreviated $p(c|x_i, \Theta_c)$ to $p(c|x_i)$ for brevity in the upcoming equations. For the update equations of our additional parameter, θ_c, we show an explicit derivation below. For simplification, we define

$$K_{i,c} = -\frac{1}{2}\left(\log|\Sigma_c| + (x_i - \mu_c)^T(\Sigma_c)^{-1}(x_i - \mu_c)\right) \quad (9)$$

Substituting equations (9) and (5) into equation (7), and dropping constant terms, our new maximization expression is

$$\sum_{c=1}^{k}\sum_{i=1}^{N} \Re\left[(\chi_l^{(i)})^* \cdot e^{j\theta_c}\right] K_{i,c} p(c|x_i) \quad (10)$$
$$= \sum_{c=1}^{k}\sum_{i=1}^{N} \left[\Re[\chi_l^{(i)}]\cos\theta_c + \Im[\chi_l^{(i)}]\sin\theta_c\right] K_{i,c} p(c|x_i)$$

Differentiating equation (10) with respect to θ_c and setting it to zero provides us with

$$\sum_{i=1}^{N}\left[-\Re[\chi_l^{(i)}]\sin\theta_c + \Im[\chi_l^{(i)}]\cos\theta_c\right] K_{i,c} p(c|x_i) = 0$$

$$\Rightarrow \quad \frac{\sin\theta_c}{\cos\theta_c} = \frac{\sum_{i=1}^{N} \Im[\chi_l^{(i)}] K_{i,c}\left[p(c|x_i)\right]}{\sum_{i=1}^{N} \Re[\chi_l^{(i)}] K_{i,c}\left[p(c|x_i)\right]}$$
$$\Rightarrow \quad \theta_c^{new} = \tan^{-1}\frac{\sum_{i=1}^{N} \Im[\chi_l^{(i)}] K_{i,c}\left[p(c|x_i)\right]}{\sum_{i=1}^{N} \Re[\chi_l^{(i)}] K_{i,c}\left[p(c|x_i)\right]} \quad (11)$$

and thus equation (11) is our update equation for θ_c^{new}. Thus, we have update equations for all of our cluster parameters.

One problem remains with this approach, however, and it is illustrated in Figure 3. If a coefficient at location x_i is opposite-phase to the nearest cluster c_1, its β_{i,c_1} value will be zero, and its "membership" in class c_1, represented by $p(c_1|x_i)$, will also be zero. Therefore, it will instead have strong membership in the next closest similar-phase class, c_2, regardless of the distance between x_i and μ_{c_2}. As a result, the covariance matrix Σ_{c_2} will expand inappropriately to include x_i, and μ_{c_2} will be biased. To prevent this effect, we alter $\beta_{i,c}$ to *disregard* phase differences for the purpose of membership calculation *only*:

$$\widetilde{\beta}_{i,c} = |\chi_l^{(i)}| \quad (12)$$

Our likelihood expression for membership calculation is now

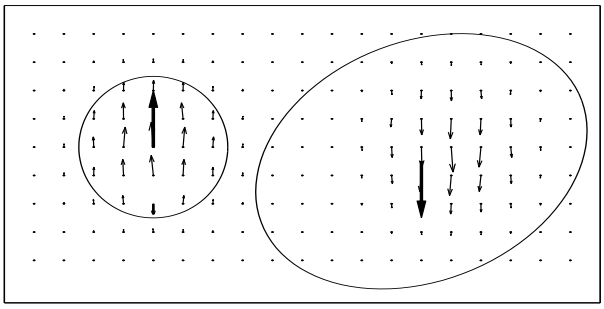


Fig. 3. A demonstration of the shortcomings of the "pure" EM procedure described above. In this two-cluster example, the highlighted coefficient near the bottom of the left cluster (c_1) actually has 100% membership in the right cluster (c_2), due to its phase. As a result, cluster c_2 expands inappropriately to accommodate this coefficient, and becomes too large and off-centered. We therefore use Equations 12-14 to ignore phase as a factor in cluster membership.

$$\widetilde{p}(x_i, \beta_{i,c}|\mu_c, \Sigma_c, \theta_c) = \frac{1}{Q_i}\left[\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_c|^{\frac{1}{2}}} e^{-\frac{1}{2}((x_i-\mu_c)^T\Sigma_c^{-1}(x_i-\mu_c)}\right]^{\widetilde{\beta}_{i,c}} \quad (13)$$

And, from Bayes' Rule, the resulting expression for the membership function is now

$$p(c|x_i) = \frac{\widetilde{p}(x_i, \beta_{i,c}|\mu_c, \Sigma_c, \theta_c)\cdot\alpha_c}{\sum_{d=1}^{k}\widetilde{p}(x_i, \beta_{i,d}|\mu_d, \Sigma_d, \theta_d)\cdot\alpha_d} \quad (14)$$

With this membership function, we ensure that our clusters remain relatively compact, although we cannot guarantee homogeneity in phase.

We initialize the parameter values of each cluster to have $\theta_c^0 = 0°$, $\alpha_c^0 = \frac{1}{k}$, and $\Sigma_c^0 = I$. Initial values for μ_c^0 are established by using a non-maximal suppression algorithm based on the magnitudes of the ILP coefficients, with a radius of suppression that decreases until k well-separated peaks are identified.

4. RESULTS

In Figure 4, we see the ILP coefficients of Figure 2 clustered with 10 coefficients per subband. Note that clusters are developed for each subband independently; there is no interaction between subbands in the clustering procedure. As with all clustering algorithms, determining the number of clusters that are required to effectively represent image features is an important aspect, and work will continue in this area. For this example, iterations continue until all cluster means

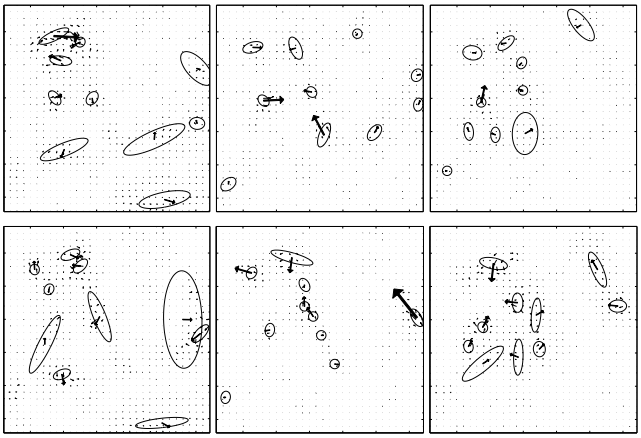


Fig. 4. Trained Gaussian clusters, 10 per subband, to represent the data shown in Figure 2. The angle of the arrow in each cluster represents the ILP direction θ_c, and the magnitude represents α_c, the relative weight of the cluster.

Fig. 5. Perceptual interpretation of example clusters in Figure 4, where the phases of each cluster are interpreted using the relationship in Figure 1 at the appropriate subband orientation.

and complex directions shift less than 0.1 samples and 0.1 radians between iterations.

Our ILP clustering technique has been used upon several aerial images with varying degrees of success. Predictably, its greatest successes occur on images with clear structural components; that is, images containing objects with dominant, well-spaced multiscale edges and ridges. As one may expect, random textures and flat areas are difficult to cluster into sparse feature representations, and ILP clusters should accordingly be used only to represent objects and areas that are defined by their shapes, rather than their textures.

We have two criteria by which we can measure the effectiveness of our clustering algorithm: perceptually, and numerically. Perceptually, we can compare cluster data with our own perception of the image by looking at the heaviest-weighted clusters in the image and associating them with image features; in Figure 5, we see examples of the 15° step edge in the upper left farm, for instance, and the dominating 135° ridge-edge in subband 5 in the right of the diagram.

Numerically, we can regenerate the ILP coefficients and compare the results with the original ILP coefficients through a mean-squared error measure; or, equivalently, we could use a normalized cross-correlation. Note that we cannot sensibly use differences between the actual image and a cluster-generated image, as small changes in the ILP can produce large changes in the original image. With such measures, we can find the minimal number of clusters required to capture a given proportion of overall ILP data.

5. CONCLUSIONS

This paper has introduced a method by which directional ILP data may be clustered with a modified Gaussian mixture. We show that this method can be used to sparsely represent major image components at a given level of scale, and the parameters of these clusters have clear perceptual meaning. By reducing image content into this sparse format, we are able to perform fast searches for various permutations and affine transformations of a desired object, in a decimated search space. We therefore plan to use these ILP clusters to build scale- and rotation-invariant representations of image objects that can be used for fast image retrieval and object recognition.

6. REFERENCES

[1] N.G. Kingsbury, "Complex wavelets for shift invariant analysis and filtering of signals," *Journal of Applied and Computational Harmonic Analysis*, no. 3, pp. 234–253, 2001.

[2] R. Anderson, N. Kingsbury, and J. Fauqueur, "Coarse level object recognition using interlevel products of complex wavelets," in *International Conference on Image Processing (ICIP)* , September 2005.

[3] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal Royal Stat. Soc., Series B*, vol. 39, no. 1, pp. 1–38, 1977.

[4] J. Blimes, "A gentle tutorial on the EM algorithm and its application to parameter estimation for Gaussian Mixture and Hidden Markov Models.," Tech. Rep., University of Berkeley, 1997.