# Perceptual Feature Detection

Naila Murray, Xavier Otazu and Maria Vanrell

*Colour in Context Group*
*Computer Vision Centre, Edifici O - Campus UAB, 08193 Bellaterra, Spain*
*naila, xotazu, maria@cvc.uab.cat*

## Abstract

Currently there exists no application-independent or general theory of feature detection. In this work, a brightness induction wavelet model (BIWaM) is extended with the long-term aim of developing a principled model for generic local feature detection. This detector, the Feature Induction Wavelet Model (FIWaM), uses the same "featureness" measure for a range of local features such as blobs, bars and corners. FIWaM is a wavelet-based computational model that attempts to use the perceptual processes involved in visual brightness induction to enhance and detect these features. The model uses two center-surround mechanisms in sequence to detect features - a Gabor-like mother wavelet followed by an explicitly-defined center-surround region mechanism. These center-surround regions are feature-specific and introduce the only variation in the detection schema between features. Preliminary results have shown that this mechanism is effective in detecting features and achieves a repeatability performance in line with current state-of-the-art detection methods.

*Keywords*: Brightness Induction, Feature Detection, Wavelet Transform.

## 1 Introduction

Feature detection has an essential role in many important computer vision tasks, including image matching and registration, object recognition and tracking and scene classification. Consequently there has been a plethora of research devoted to developing efficient and effective feature detection techniques. Currently, state-of-the-art detectors use very different methods and, as a result, their performance differs widely depending on the data sets they are used to analyse. To date, there exists no application-independent or general theory of feature detection. Therefore, determining which feature detector to use on any specific application tends to require *a priori* information about the data set, and a subjective judgement on the most suitable method for feature detection.

This paper extends the perceptual processes present in a low-level human visual system (HVS) model of brightness induction, (BIWaM) [7], with the long-term aim of developing a principled model for generic local feature detection. Several successful detectors [4, 6] have been modelled using "biologically plausible architecture" [3] related to the HVS with much success. The motivation for using the BIWaM is to incorporate many relevant attributes of the HVS with the aim of combining the advantages of detectors which use these attributes separately.

## 2 Related Research

As mentioned in the previous section, biologically-inspired frameworks have been employed successfully in local feature detection. Lowe's SIFT algo-

rithm [4] and Mikolajczyk & Schmid's Hessian-Affine and Harris-Affine algorithms [6] are all based on multi-scale image decomposition. In addition, the Difference of Gaussian (DoG) kernel used in the SIFT scale decomposition has a center-surround profile and thus can be thought of as a centre-surround response mechanism.

Collins & Ge [2] employ the multi-scale concept, as well as an explicit centre-surround mechanism, for feature extraction. Centre and surround regions were defined using a Laplacian of Gaussian kernel. The positive (circular) region of the kernel corresponds to the central local region while the negative (annular) region of the kernel corresponds to the surround local region. For the center, the kernel is used to weight values around a location in a Gaussian fashion. A distance measure is calculated for the centre and surround regions and compared to that of neighbouring pixels. As with SIFT, local extrema are selected as candidate features.

Agrawal et al. [1] adopt a similar approach to that of Collins & Ge [2], but the DoG kernel is simplified to a Difference of Boxes (DoB) kernel. Also, bi-level center and surround boxes are used, i.e. with values of 1 or -1, in order to enable an extremely simple and fast computation. Finally, extrema are extracted at different scales not by blurring and down-sampling the image but by changing the scale of the kernels.

# 3 The BIWaM Model

The BIWaM [7] modifies a visual stimulus in order to reproduce the brightness induction performed by the HVS. Brightness induction refers collectively to

- brightness assimilation, where the brightness of a visual target (considered the center region) becomes more similar to that of the surrounding region, and
- brightness contrast, where the brightness of a visual target becomes less similar to that of the surrounding region.

The model is based on three main assumptions, derived from known psychophysical phenomena:

1. Induction is higher between features of similar **spatial scale**. Because image features are isolated by spatial scale in wavelet planes, this is achieved using the image decomposition. As Figure 1(a) illustrates, induction is strongest between features within one octave of each other in scale space.
2. Induction is higher between features of similar **spatial orientation**. Inhibition is strongest when orientations are identical, while facilitation is strongest when orientations are orthogonal (see Figure 1(b)). This is also inherent in the wavelet decomposition.
3. Induction is modulated by the **stimulus-surround relative contrast**. For increasing surround contrast there is increasing inhibition and vice-versa, as can be seen in Figure 1(c).

## 3.1 The Wavelet Decomposition

The wavelet decomposition of the image is a key point of the model. Images are decomposed into a series of new images (wavelet planes) with respect to spatial scale $s$ and orientation $o$ (vertical, horizontal and diagonal), which is inspired by parvocellular spatial frequency channels and cortical orientation-selective receptive fields in the HVS. The wavelet planes, $w_s^h$, $w_s^v$ and $w_s^d$, contain the response of the image intensities at that orientation to the wavelet kernel corresponding to the scale, $s$.

The image, $I$, is reconstructed as:

$$I = U_{s=1} \tag{1}$$

where $U_s$ is the $s$-th element of a recursive series of images

$$U_s = (U_{s+1} \uparrow 2) + d_{s+1} \tag{2}$$

(a) Assimilation is greatest when spatial scales are identical.

(b) The relative orientations must be identical for maximum induction.

(c) The induction effect increases with an increase in surround contrast.

Figure 1: The assumptions of the model.

and d is the sum of the oriented wavelet planes:

$$d_s = \sum_{o=v,h,d} w_s^o, \tag{3}$$

where $\uparrow 2$ denotes up-sampling by a factor of 2.

### 3.2 Construction of the Perceived Image

Equation 2 describes the reconstruction of the original image from the wavelet decomposition. The perceived image is obtained from this reconstruction with the simple introduction of a weighting function $\alpha$, which is designed according to the assumptions of the induction model. The modified image recovery defined by Equation 4:

$$I' = U_1^\alpha = (U_{s+1}^\alpha \uparrow 2) + \alpha \cdot d_{s+1} \tag{4}$$

introduces $\alpha$, thereby generating the perceived image, $I'$.

### 3.3 The $\alpha$ Weighting Function

The weighting function $\alpha$ can be seen as a generalisation of the psychophysically-determined Contrast Sensitivity Function, (CSF), $C_d$. It has been shown that the HVS is very sensitive to mid-range frequencies, and to a lesser extent to low frequencies. It is important to note that frequencies are relative to the distance, $d$, between the viewer and the visual stimulus. This concept of viewer distance was incorporated into the definition of $C_d$ (see Appendix A of [7]). The weighting function is based on this CSF but has been modified to introduce the effect of surround contrast and is defined as

$$\alpha(s, z_{ctr}) = z_{ctr} \cdot C_d(s) + C_{min} \tag{5}$$

The $z_{ctr}$ term defined by

$$z_{ctr} = \frac{r^2}{1 + r^2} \tag{6}$$

where $r = \frac{\sigma_{cen}}{\sigma_{sur}}$, introduces relative contrast energy implicitly. The standard deviation, $\sigma$, of a region is used as a measure of its self contrast. Therefore the ratio $r$ is the relative contrast energy of the center and surround regions. The $r$ term is dependent on orientation $o$. To avoid null $\alpha$ values, the $C_{min}$ term was introduced.

## 4 Perceptual Feature Detection

The brightness induction framework of the BIWaM can be modified to accomplish feature induction by substituting $\alpha$ for a suitably designed weighting function. As such we present what we term the Feature Induction Wavelet Model (FIWaM), with a new weighting function $\beta$, that modifies Equation 4 as follows:

$$I' = U_1^\beta = (U_{s+1}^\beta \uparrow 2) + \beta \cdot d_{s+1} \tag{7}$$

It is apparent from Equation 7 that the modified image recovery is identical to that of the BIWaM except for the new weighting function, $\beta$.

However, to detect features using the FIWaM, $\beta$ is used as a feature detection measure, rather than a weighting function and is defined using two hypotheses:

1. Features are present within a bounded range of scale space.

2. If a stimulus region's response to a feature's characteristic shape is appropriately large, the stimulus region contains that feature.

With these hypotheses in mind, we define $\beta$ as:

$$\beta(s, z_{ctr}) = \gamma \cdot z_{ctr} \cdot C_{det}(s). \quad (8)$$

The new CSF, $C_{det}(s)$, is an ideal band-pass filter that bounds the range of scale space in which features are detected.

The $z_{ctr}$ term is defined as previously. However, the center and surround regions are now defined differently for each feature, in order to reflect the feature's characteristic shape. Therefore, $z_{ctr}$ measures the stimulus's response to a specific feature's shape. The median contrast energy term, $\gamma$:

$$\gamma = \mid median_{cen} - median_{sur} \mid \quad (9)$$

is the difference between the median intensity values of the stimulus and stimulus-surround regions and quantifies the strength of the wavelet response of the stimulus. Together, $z_{ctr}$ and $\gamma$ measure the type and the strength of a detected feature, respectively. Therefore, $\beta$ constitutes a "featureness" measure, that is, the degree to which a stimulus corresponds to a feature.

### 4.1 Characterisation of Feature Shapes

We have investigated detection with respect to four features - blobs, bars, corners and terminators. These regions have a size that corresponds to the minimum size of interest of the feature. They also reflect the appearance of the feature decompositions in the wavelet plane, as shown in Table 1.

| Feature | Feature representation | Wavelet plane representation | Center | Surround |
|---|---|---|---|---|
| Blob |  |  |  |  |
| Bar |  |  |  |  |
| Corner |  |  |  |  |
| Terminator |  |  |  |  |

Table 1: Wavelet decompositions of features along with their center and surround region definitions.

### 4.2 Feature Selection

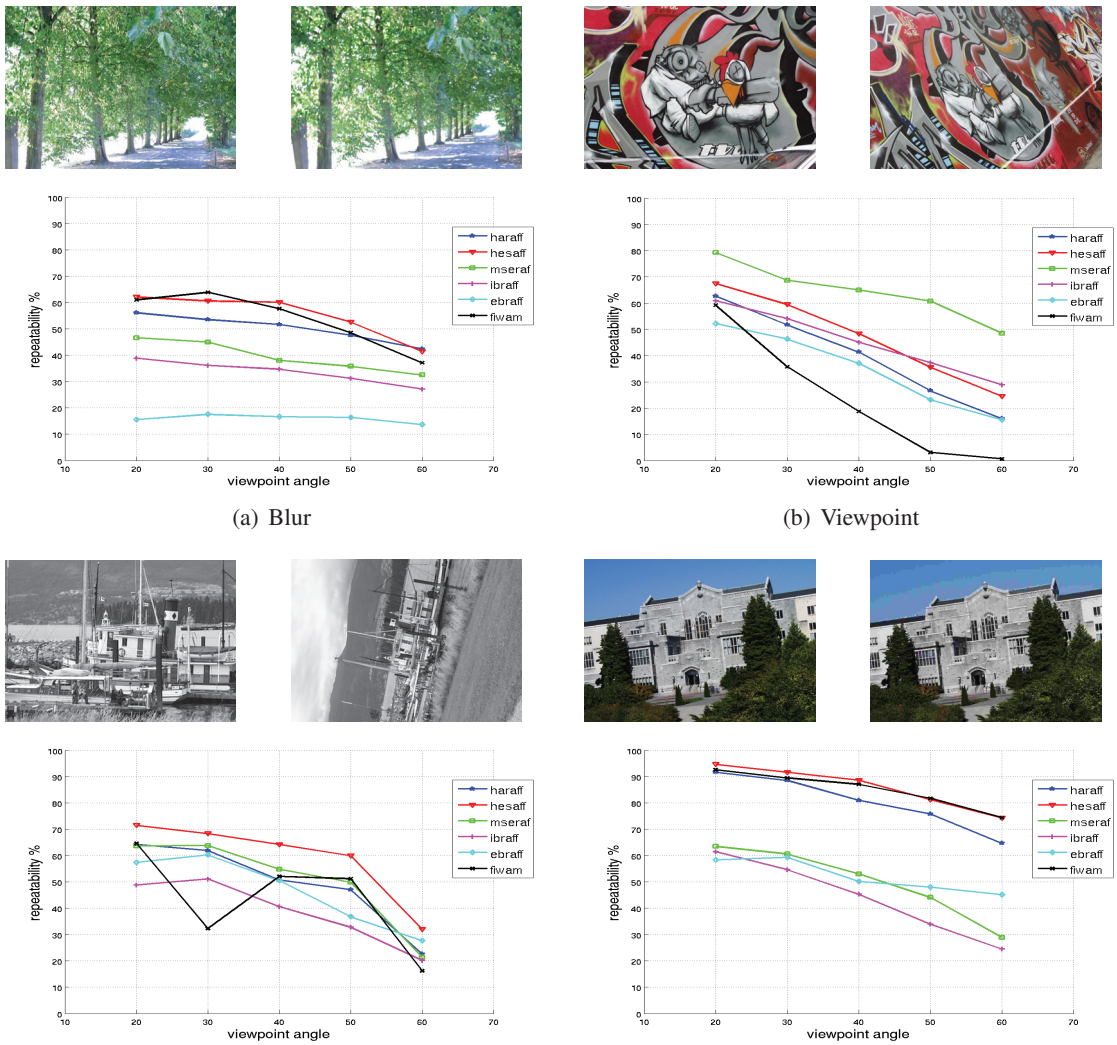
To select a stimulus region as a feature, the feature must have a $\beta$ value that is a local extremum in $(x, y, \sigma)$-space, where $\sigma$ signifies scale-space. In addition, $\beta$ must be over a certain threshold to ensure the feature is strong, that is, more repeatable.

For each image, detection is performed separately for the 4 types of features described, in all possible orientations. The aggregation of the features from these separate detections comprises the final feature set for an image.

One sometimes finds that different features, for example both a blob and a bar, are detected at the same spatial location. In such cases, the different featureness responses are compared and the feature with the highest featureness response is selected.

## 5 Experiments

To asses the detector's performance, the repeatability of the detected features was tested using the experimental framework constructed by Mikolajczyk et al. [5]. In this experiment, feature detection is performed on a sequence of images of the same scene. One of these images is considered to be the reference image and there exists a known homographic relationship between the reference image and the other images in the sequence. Therefore, for an image in the sequence, the regions of the




(a) Blur (b) Viewpoint

(c) Zoom & rotation (d) JPEG compression

Figure 2: Repeatability for different transformations: The FIWaM plot is shown in black. Examples of data set images are shown above the repeatability plots.

scene contained in both the image and the reference image are known. A feature is considered to be repeated if it is found in a region common to both images and it has been detected in both images. Additionally, the spatial overlap of the regions defined by the features must be greater than a user-defined threshold. An overlap threshold of 40% is typical and was used here. The experiment was conducted on 4 sequences of six images with homographic variances with respect to Gaussian blurring, viewpoint, zoom & rotation and JPEG compression. The transformation increases in severity along the sequence of images.

The repeatability results are shown in Figure 2. For comparison, data for five state of the art feature detectors have been included: Harris-Affine (HARAFF), Hessian-Affine (HESAFF), Maximally Stable External Region (MSERAF), Intensity Extrema-Based Region (IBRAFF) and Edge-Based Region (EBRAFF) [5].

One should note that repeatability is not the only criteria available for evaluating a detector's performance. However, repeatability has shown itself to be a good general indicator of performance, and so it was used here.

### 5.1 Discussion

It is evident that the FIWaM detector performs comparably with respect to state-of-the-art detectors, except in cases of severe affine transformation, such as in the graffiti sequence (Figure 2(b)). This is unsurprising given that the FIWaM detector has coarse affine estimation. Firstly, the wavelet's scale decomposition may be too coarse for accurate scale localisation of features. For a typical image decomposition of 7 octaves, 5 octaves are analysed due to the nature of the CSF. This means that features have only 5 possible scales. Secondly, the elliptical estimation is derived from the shape of the center regions, not the shape of the stimulus itself, resulting in an imprecise estimation.

## 6 Conclusions

In this paper, a brightness induction wavelet model (BIWaM) was extended with the long-term aim of developing a principled model for generic local feature detection. It has been shown that the brightness induction model can be modified successfully to create an effective local feature detector.

However, there are many avenues for further exploration. Most promising is improving the center and surround regions for several features so that they more closely resemble the appearance of these features in the wavelet decomposition. This would improve the problem of affine variance, as would using a more refined scale-space decomposition.

In addition, in this work, there has been no discussion on incorporating colour information. However, there exists a straight-forward extension of the BIWaM to colour, namely the Colour Induction Wavelet Model (CIWaM). Incorporating colour information may allow the detection of features in channels other than intensity, such as in the opponent colour space.

## References

[1] Motilal Agrawal, Kurt Konolige, and Morten Rufus Blas. Censure: Center surround extremas for realtime feature detection and matching. In David A. Forsyth, Philip H. S. Torr, and Andrew Zisserman, editors, *ECCV (4)*, volume 5305 of *Lecture Notes in Computer Science*, pages 102–115. Springer, 2008.

[2] Robert T. Collins and Weina Ge. Csdd features: Center-surround distribution distance for feature extraction and matching. In *ECCV '08: Proceedings of the 10th European Conference on Computer Vision*, pages 140–153, Berlin, Heidelberg, 2008. Springer-Verlag.

[3] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(11):1254–1259, 1998.

[4] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[5] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *Int. J. Comput. Vision*, 65(1-2):43–72, 2005.

[6] Krystian Mikolajczyk and Cordelia Schmid. Scale & affine invariant interest point detectors. *Int. J. Comput. Vision*, 60(1):63–86, 2004.

[7] X Otazu, M Vanrell, and C.A Párraga. Multiresolution wavelet framework models brightness induction effects. *Vision Research*, 48:733–751, February 2008.