Spatio-Chromatic Decorrelation by Shift-Invariant Filtering

Matthew Brown, Sabine Süsstrunk and Pascal Fua
School of Computing and Communication Sciences,
École Polytechnique Fédérale de Lausanne (EPFL).
{matthew.brown,sabine.susstrunk,fua}@epfl.ch

Abstract

In this paper we derive convolutional filters for colour image whitening and decorrelation. Whilst whitening can be achieved via eigendecomposition of the image patch covariance, this operation is neither efficient nor biologically plausible. Given the shift invariance of image statistics, the covariance matrix contains repeated information which can be eliminated by solving directly for a per pixel linear operation (convolution). We formulate decorrelation as a shift and rotation invariant filtering operation and solve directly for the filter shape via non-linear least squares. This results in opponent-colour lateral inhibition filters which resemble those found in the human visual system. We also note the similarity of these filters to current interest point detectors, and perform an experimental evaluation of their use in this context.

1. Introduction

According to the efficient coding hypothesis, the goal of the visual system should be to encode the information presented at the retina with as little redundancy as possible. From the signal processing point of view, the first step in removing redundancy is decorrelation, which removes the second order dependencies in the signal. This principle was explored in the context of trichromatic vision by Buchsbaum[3] and later Ruderman [14], who found that linear decorrelation of LMS cone responses at a point matches the opponent colour coding in the human visual system. Spatial decorrelation is also evident in human vision; lateral inhibition operations which decorrelate spatially result in the well known visual illusion of Mach Bands [12].

Similarly, most successful techniques for interest point detection in computer vision rely directly or indirectly on decorrelation. For example, the commonly-used difference of Gaussian detector [9] is in fact the linear whitening filter for greyscale images. Similarly, the Harris corner detector [7] finds points where the local sum-square difference function, which is inversely related to the autocorrelation, is peaked in all directions. Comparitively little work has gone into exploiting colour, although [15] provides a generalisation of Harris corners to colour images, and [5] derives a colour stable region detector. Using grayscale-only detectors discards potentially discriminating information in the chromaticity channels, and in the extreme case of isoluminant images, all greyscale detectors will fail in the same way as naive grey conversion algorithms do [6].

Though spatio-chromatic decorrelation has been explored in the context of human vision [14] and signal compression [4], the convolutional filters to effect it were not made explicit in these works. Matrix decomposition techniques such as PCA or ZCA are often used to whiten colour images [11, 8]. However, these formulations ignore shift and rotation invariance, leading to a redundant parameterisation. This results in lower fidelity solutions and the risk of overfitting.

In this work we formulate spatio-chromatic decorrelation as a shift invariant linear operation (convolution), and solve directly for the filter shape that effects it. This provides an efficient way to decorrelate colour images. We also show an application of these filters to colour interest point detection.

2. Decorrelation and Shift Invariance

A standard approach to decorrelation/whitening is to diagonalise the covariance matrix of the signal

$$\mathbf{C}' = \mathbf{W}\mathbf{C}\mathbf{W}^T = \mathbf{I}, \tag{1}$$

where $\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i^T$ is the covariance of the centred data $\mathbf{x}$, and $\mathbf{C}'$ is the covariance of $\mathbf{x}'$ after applying the whitening transform $\mathbf{x}' = \mathbf{W}\mathbf{x}$. There are multiple solutions for $\mathbf{W}$; for example, whitening via PCA would project using $\mathbf{W} = \mathbf{\Sigma}^{-1/2}\mathbf{U}^T$. The symmetrical solution

$$\mathbf{W} = \mathbf{U}\mathbf{\Sigma}^{-1/2}\mathbf{U}^T \tag{2}$$

preserves the phase of the input and is called ZCA [1] ($\mathbf{U}$ and $\mathbf{\Sigma}$ contain the eigenvectors and eigenvalues of $\mathbf{C}$). Note
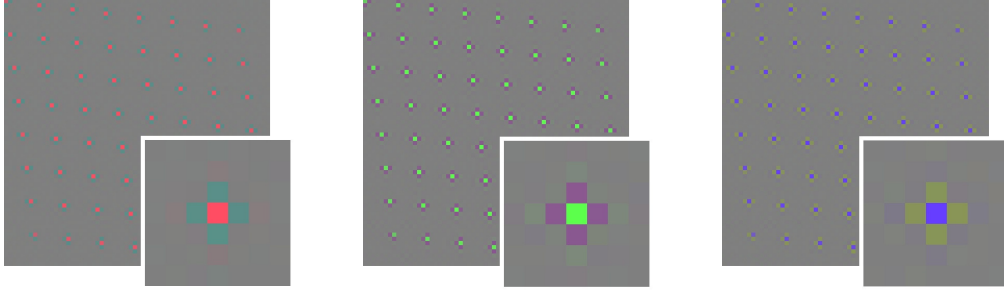
Figure 1: ZCA for colour images. The rows/columns of the symmetric whitening matrix $\mathbf{W}$ are shifted versions of each other (large images), so the whitening transform effectively consists of a convolution with three colour filters (inset images).

that one can also decorrelate without whitening by multiplication by $\mathbf{U}^T$. The results of applying ZCA to a colour image covariance matrix sampled from 10^6 pixels in 1000 images [10] are shown in Figure 1. The three larger images visualise the rows/columns of $\mathbf{W}$. As can be seen, the columns are all shifted versions of each other, so that multiplication by the whitening matrix $\mathbf{W}$ is effectively a convolution with the 3 colour filters shown in the inset images. This structure is not explicitly enforced, but arises because of the shift invariance of image statistics. This motivated us to explicitly look for a shift-invariant linear operation (i.e., a convolution), that whitens colour images.

In addition to shift invariance, images also have rotation and scale invariance. The scale invariance of image statistics is well known [13], and is observed for example in the power law distribution of amplitude spectra, $A(\omega) \propto 1/\omega$. Rotation invariance may not be exactly present in all cases, for example, human authored images of man-made scenes have more energy in the horizontal and vertical directions [16]. However, human images of natural scenes are almost rotation invariant, and rotation invariance is also a desirable property in many matching applications, so we will enforce it here also.

Given shift, rotation and scale invariance, the second order statistics may be encapsulated in a 1-dimensional autocorrelation function, which measures the similarity of pixels in any direction, at any scale and at any position. Formulating our whitening filters as 1-dimensional functions will allow us to use a reduced parametrisation, helping to prevent overfitting, and we will also be able to handle longer range correlations than possible with PCA/ZCA (e.g. computing ZCA on 64x64 colour image patches requires factorisation of a 36,864×36,864 matrix).

2.1. Spatial Decorrelation and DOG

We can whiten a shift-invariant signal $I(x)$ by convolving with a filter $h(x)$ such that the autocorrelation of the output equals the dirac delta function. We start by computing the image autocorrelation function, by sampling the image in a straight line at random positions, orientations and scales

$$r_I(\tau) = \frac{1}{N} \sum_{k=1}^{N} I_k(x) I_k(x + \tau). \qquad (3)$$

$I_k(x)$ represents the image sampled at a random position, orientation and scale. This is represented for a greyscale image in Figure 2, leftmost plot. We then find the inverse filter $h(x)$ which satisfies

$$r_I(x) * r_h(x) = \delta(x) \qquad (4)$$

where $\delta(x)$ is the unit impulse, and $r_h(x) = h(x) * h(x)$ is the autocorrelation of the filter. Equivalently

$$P_I(\omega) P_H(\omega) = 1 \qquad (5)$$

where $P_I(\omega)$ is the power spectrum of the signal (Fourier transform of the autocorrelation) and $P_H(\omega)$ is the power spectrum of the filter. One could compute $H(\omega)$ and thus $h(x)$ by inverting the square root of the power spectrum, with suitable priors on the high frequency components of $h(x)$. Instead, we choose to solve Equation 4 directly in the spatial domain, solving for the whitening filter with minimum squared intensity error for a smoothed pulse output

$$h^*(x) = \min_{h(x)} \sum_x |r_I(x) * (h(x) * h(x)) - p(x)|^2$$
$$+ \lambda \sum_x (\rho(x) h(x))^2 . \qquad (6)$$

We relax the requirement of complete decorrelation by setting $p(x) = g(x; 0, \sigma^2)$, with a small σ of around 4 pixels. $h(x)$ is constrained to be symmetric i.e., $h(-x) = h(x)$, and we apply a weighting $\rho(x)$ that encourages $h(x)$ to fall to zero as x becomes large (we have used $\rho(x) = (1 + (2x/n_h)^2)$ where n_h is the size of the filter). Equation 6 is solved using standard non-linear least squares solvers and converges well from any random initialisation.
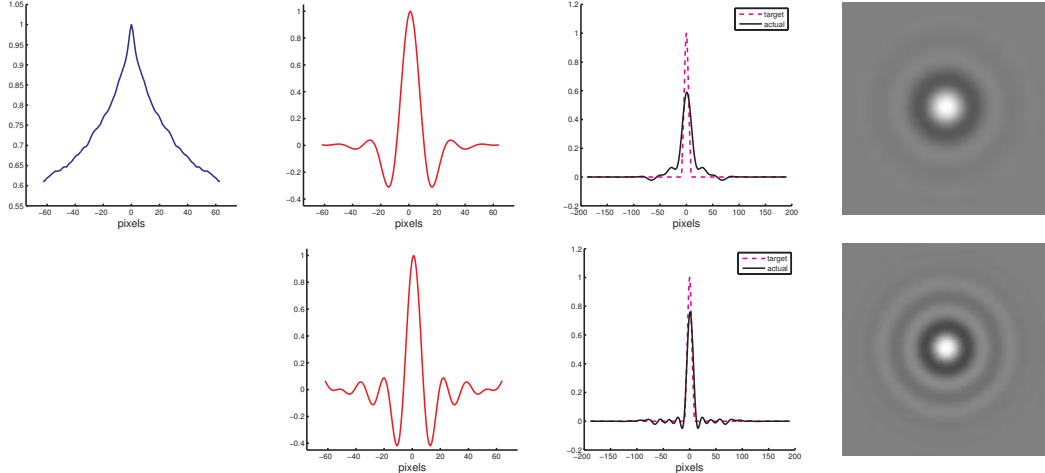
Figure 2: Greyscale decorrelation. The difference of Gaussian filter is an effective decorrelation filter for greyscale images. We solve for a whitening filter (centre-left) that converts the long range image correlations (top-left) to leave only residual correlations with nearby pixels (centre-right). Adding a prior on the energy and spatial extent of the result gives a DOG like function (top-right), with less ringing than the unsmoothed version (bottom-right).

The solution is visualised in Figure 2. The image is convolved with a symmetric filter (centre-left column), so that the target output autocorrelation is a smoothed pulse (centre-right column). The solution is well modelled by a difference of Gaussian function. This result provides some justification as to why the popular difference of Gaussian operation (used for example in SIFT [9]) might be a good one in early image understanding, e.g., interest point detectors – it decorrelates a greyscale input image.

2.2. Chromatic Decorrelation and Opponent Colours

The RGB colour channels are also strongly correlated, with overall changes in intensity affecting each channel almost equally and making up the majority of the signal energy. The eigenvectors of the colour correlation matrix also have strong connection to human vision, as pointed out by Buchsbaum [3]. If we represent the colour information using the theoretical LMS (long, medium, short wavelength) cone responses, the principal components correspond to luminance ($\approx$95% energy), and the opponent chrominance channels of blue-yellow and red-green (see Figure 3, left). The eigenvectors of sRGB images are slightly different (see Figure 3). A smaller, but still large fraction (80%) of the energy is in an achromatic channel, with colour differences of red-blue and green-purple making up the remaining decorrelated channels. These results were computed using 1000+ calibrated images from the McGill Colour Image Database [10].

The rightmost plot in Figure 3 shows the zero-phase whitening matrix $\mathbf{W} = \mathbf{U}\boldsymbol{\Sigma}^{-1/2}\mathbf{U}^T$ for the sRGB colour

channels. Note that there are multiple whitening transforms, corresponding to arbitrary rotations of this matrix.

2.3. Combining Spatial and Chromatic Decorrelation

After having found whitening operations for spatial and chromatic dimensions separately, it seems natural to investigate the joint objective, i.e., to find a decorrelating filter for both space and chromaticity togther. In this case, the whitening is achieved via convolution with a matrix function

$$\mathbf{I}'(x) = \mathbf{H}(x) * \mathbf{I}(x) \qquad (7)$$

so that each whitened output in $\mathbf{I}' = \begin{bmatrix} r' & g' & b' \end{bmatrix}$ is the convolution of 3 channel colour filter (one row of $\mathbf{H}(x)$) with the input ($*$ denotes matrix convolution), i.e.,

$$\begin{bmatrix} r'(x) \\ g'(x) \\ b'(x) \end{bmatrix} = \begin{bmatrix} h_{rr}(x) & h_{rg}(x) & h_{rb}(x) \\ h_{gr}(x) & h_{gg}(x) & h_{gb}(x) \\ h_{br}(x) & h_{bg}(x) & h_{bb}(x) \end{bmatrix} * \begin{bmatrix} r(x) \\ g(x) \\ b(x) \end{bmatrix}. \quad (8)$$

The whitened image autocorrelation function is

$$\mathbf{R}_{\mathbf{I}'}(\tau) = \mathbb{E}\left(\mathbf{I}'(x)\mathbf{I}'(x+\tau)^T \right) \qquad (9)$$

$$= \mathbb{E}\left(\mathbf{H}(x) * \mathbf{I}(x) . \mathbf{I}(x+\tau)^T * \mathbf{H}(x)^T \right) \quad (10)$$

$$= \mathbf{H}(x) * \mathbf{R}_{\mathbf{I}}(\tau) * \mathbf{H}(x)^T \qquad (11)$$

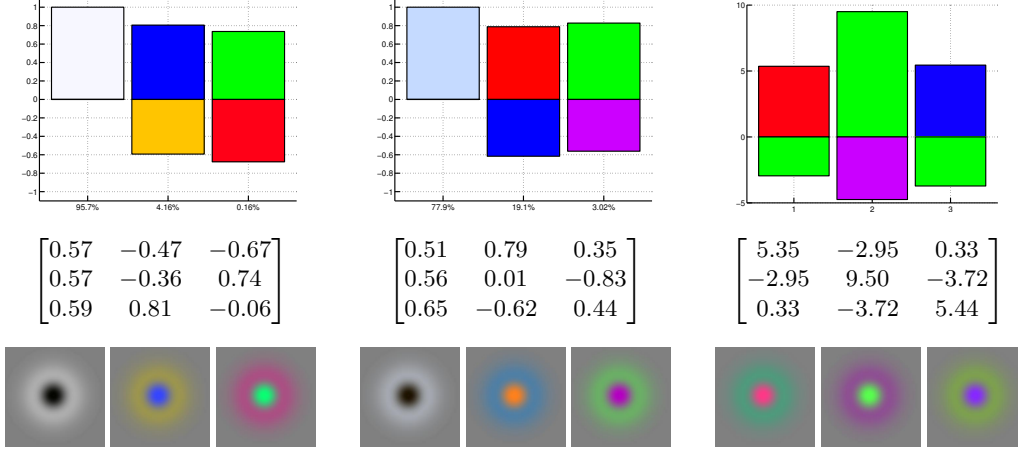Thus for whitening of the output we require

Figure 3: Chromatic Decorrelation. Left: PCA of log-LMS cone responses. Centre: PCA of RGB pixel values. Right: ZCA (symmetric whitening matrix) for RGB pixel values. The lower row shows the combination of the colour space transform with a spatial difference of Gaussian function.

$$
\begin{bmatrix} h_{rr} & h_{rg} & h_{rb} \\ h_{gr} & h_{gg} & h_{gb} \\ h_{br} & h_{bg} & h_{bb} \end{bmatrix} * \begin{bmatrix} r_{I_{rr}} & r_{I_{rg}} & r_{I_{rb}} \\ r_{I_{gr}} & r_{I_{gg}} & r_{I_{gb}} \\ r_{I_{br}} & r_{I_{bg}} & r_{I_{bb}} \end{bmatrix} * \begin{bmatrix} h_{rr} & h_{rg} & h_{rb} \\ h_{gr} & h_{gg} & h_{gb} \\ h_{br} & h_{bg} & h_{bb} \end{bmatrix}^{T}
$$

$$
= \begin{bmatrix} \delta(x) & 0 & 0 \\ 0 & \delta(x) & 0 \\ 0 & 0 & \delta(x) \end{bmatrix} \quad (12)
$$

where each quantity is a function of x, but this has been dropped for clarity i.e., $r_{I_{rr}} \equiv r_{I_{rr}}(x)$ and $h_{rr} \equiv h_{rr}(x)$. Note that there are actually only 6 colour correlation functions due to symmetry, i.e., the matrix $\mathbf{R}_I(\tau)$ is symmetric. This could be written more simply as

$$
\mathbf{H}(x) * \mathbf{R_I}(x) * \mathbf{H}(x)^{T} = \mathbf{I} \quad (13)
$$

where $\mathbf{I}$ is the 3×3 identity matrix. This objective is analogous to PCA/ZCA (Equation 1), but without the redundant parameters as it explicitly takes into account the shift invariance. Though Equation 13 expresses strict equality, in practice we solve using a least squares objective for smoothed pulses, with a prior penalising the overall filter energy as in Equation 6.

The results are shown in Figure 4. We have shown the results with a weak prior (lower 3 rows), and a strong prior encouraging the filter to perform the whitening with most of it's energy concentrated towards the origin (top 3 rows). The filters again take on a difference of Gaussian form reminiscent of 2.1, with colour differences similar to 2.2. Interestingly, the difference of Gaussians for each filter are all at approximately the same scale, suggesting that they are actually just DOG filters in a transformed colour space. We have verified this result empirically by plotting the difference of Gaussian function in a ZCA whitened colour space in Figure 3 (right column).

Since the whitening transform derived above alters the amount of energy in the luminance/colour components, we experimented with the orthogonal colour decorrelation as well. We constrained the first filter to be pure luminance, and ran the same optimisation (see Figure 5). The results have a similar form to Figure 3 central column – i.e. the spatio-chromatic decorrelating filters are again DOG filters in a decorrelated colour space. This agrees with the result found by Rudermann [14] that the spatio-chromatic decorrelation is a composition of independent chromatic and spatial decorrelations. It also allows the filters to be implemented very efficiently (by a linear transform of the colour channels, followed by seperable convolution and subtraction).

3. Colour Interest Points

As noted previously, grayscale decorrelation is an effective operation for interest point detection. In the following section we test our colour decorrelation filters as colour interest point detectors. Following Lowe [9] we find interest points where the DOG function is maximised in scale and space. After applying the linear transforms introduced in the previous section, we compute a Laplacian pyramid with 3 scales per octave and a maximum of 8 octaves. Extrema are localised by finding maxima in a $3 \times 3 \times 3$ region of scale-space.

In the remainder of this section, we will refer to the results obtained using the whitening filters of Figure 4 as **wdog** and to those obtained using the opponent colour decorrelation filters of Figure 5 as **odog**. For comparison
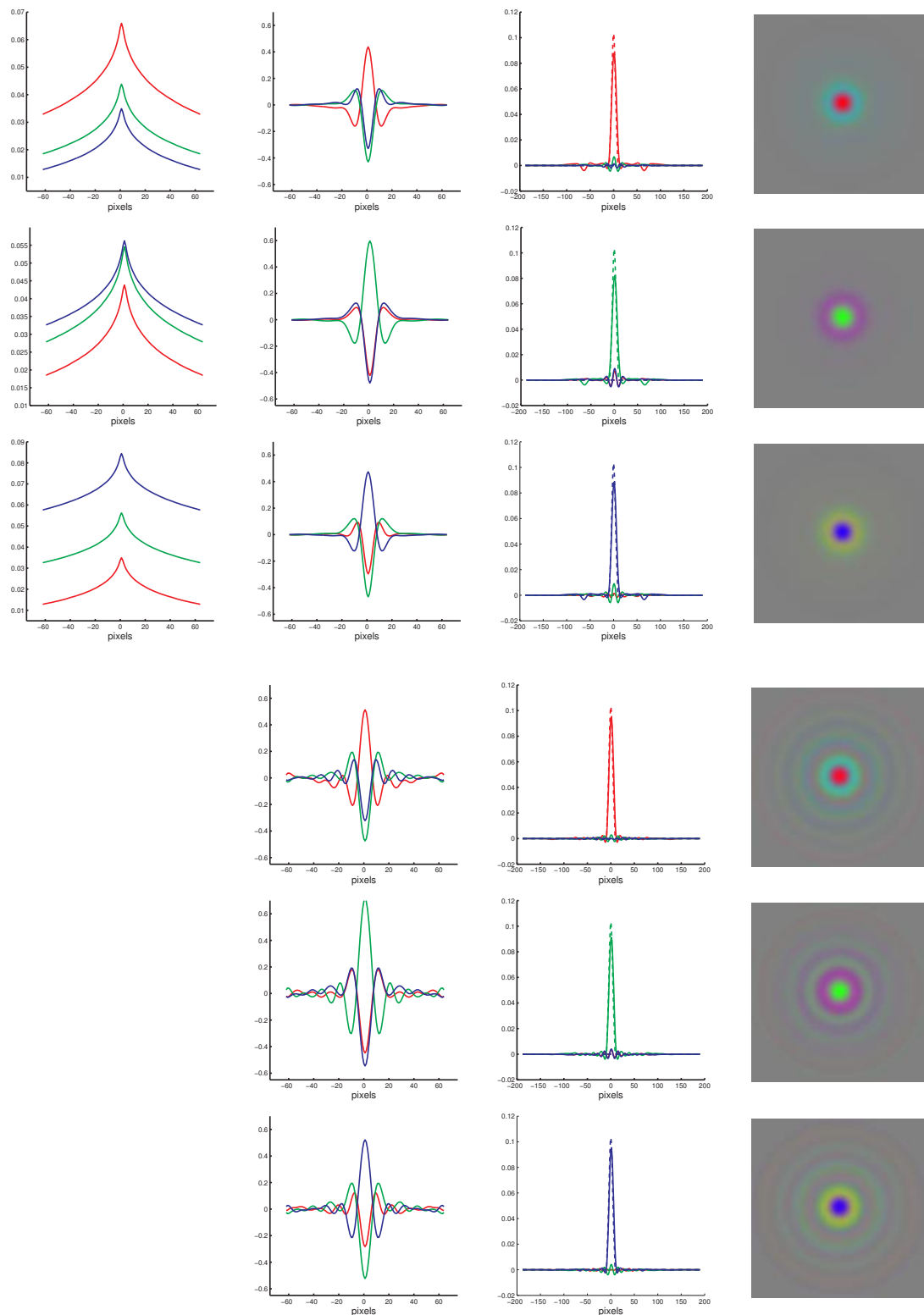
Figure 4: Colour image whitening filters (**wdog**). The image autocorrelations (left) are whitened by convolution with three colour filters (centre-left, right) to give a new signal with small residual spatio-chromatic correlations (centre-right). Applying a reduction prior to encourage parsimonious solutions results in DOG-like filters with less ringing (top three rows) compared to solving without the prior (lower three rows).
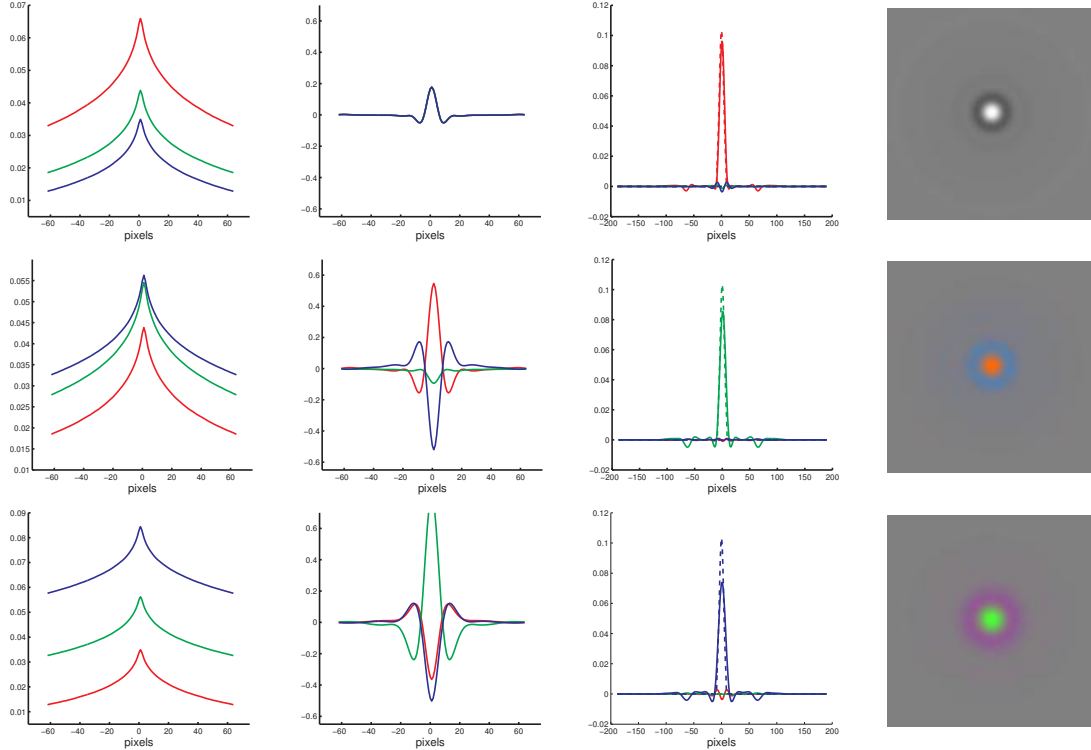
13

Figure 5: Opponent colour decorrelation filters (**odog**).

purposes, we also compute equivalent results using a standard DOG detector on a greyscale image (**dog**) and on each RGB band separately (**rgb**), the luminance channel only (**w-b**), the red-blue channel only (**r-b**), and the green-purple channel only (**g-p**). We also plot results obtained by choosing points at random (**rand**) to provide a baseline.

We compute repeatability statistics for our interest point designs using the Mikolajczyk[1] dataset. We eliminated two classes –"boat" (greyscale) and "ubc" (variable jpeg compression) – deemed irrelevant for this comparison. Since repeatability increases as more points are added to (reaching 100% when every image point is detected), we fix the number of interest points extracted from each image to 500 using a non-maximal suppression algorithm [2]. We plot the cumulative repeatability curves for epsilon in the range 0-10 pixels. In the multi-band cases, we treat all detections together in the non-maximal suppression step to avoid repeated detections.

The results are shown in Figure 6. We see that there is no clear winner in terms of repeatability across all datasets – for example, green-purple interest points give the best results on "bark" with red-blue performing poorly, yet on the "wall" dataset this performance is reversed. In half of the cases greyscale interest points give the best results, whilst in the other half colour interest points give better repeatability. However, repeatability does not tell the full story, as a key benefit of colour interest points is simply the ability to find more detections from a given size of image. In this case the question is: "how much extra information is provided by the colour channels?". To test this, we re-ran the experiments, extracting equal numbers of points in each colour band, and eliminated any duplicate detections between the bands. Duplicate detections were defined as points falling within $\epsilon = 4$ pixels of another interest point. The results are shown in Figure 7. We compare the utility of adding extra colour bands for **rgb** and **odog** points: **rgb1**: red only, **rgb2**: red+green, **rgb2**:red+green+blue, **odog1**: luminance only, **odog2**: luminance + red-blue, **odog3**: luminance + red-blue + green-purple. The striking result here is that adding extra interest points in the **odog** decorrelated colour channels is much more worthwhile than adding extra interest points in the independent red, green and blue channels. This can be explained by the fact that the RGB channels are highly correlated, so there is a much larger chance of repeat detections between the channels.

4. Conclusions

We have demonstrated a framework for explicit optimisation of shift-invariant filters for spatio-chromatic image

---
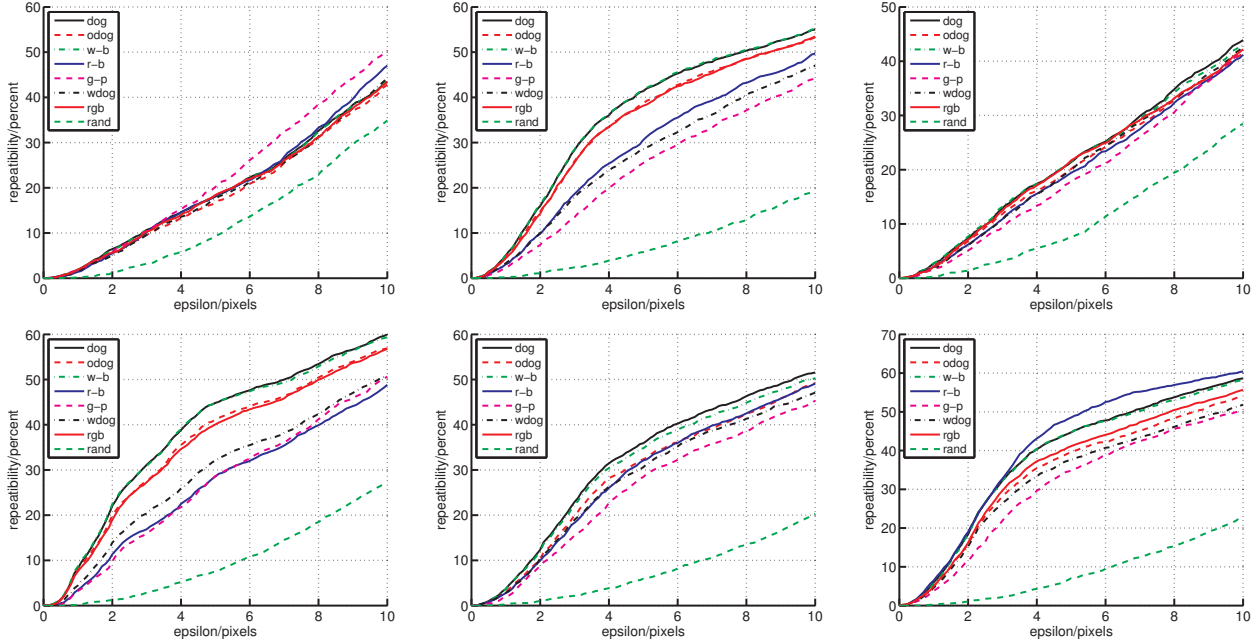
[1]http://www.robots.ox.ac.uk/∼vgg/research/affine/

Figure 6: Repeatability on the Mikolajyzck dataset (left-right, top-bottom: bark, bikes, graffiti, leuven, trees, wall). **odog**: orthogonal colour decorrelation filters, **wdog**: colour whitening filters, **dog**: greyscale DOG, **rgb**: DOG on each of RGB channels, **w-b**: luminance only, **r-b**: red-blue, **g-p**: green-purple, **rand**: uniform random distributed points. Colour DOG detectors give equal or better results to greyscale in half of the cases.
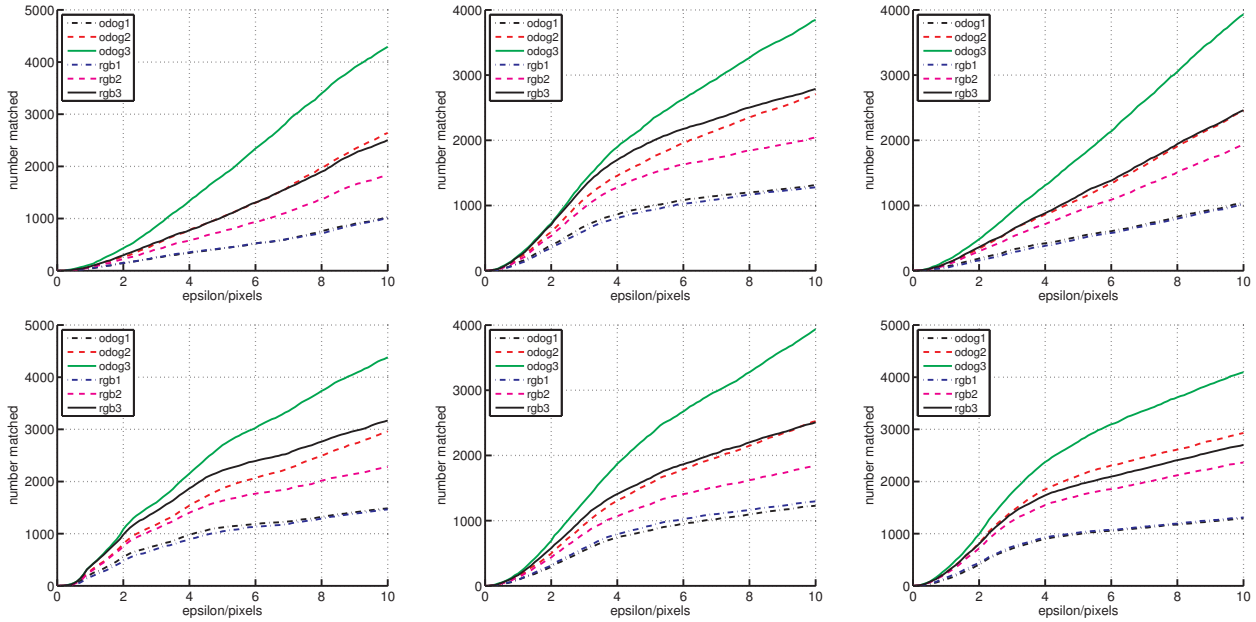

Figure 7: Number of independent matches on the Mikolajyzck dataset (left-right, top-bottom: bark, bikes, graffiti, leuven, trees, wall). **rgb1**: red only, **rgb2**: red+green, **rgb2**: red+green+blue, **odog1**: luminance only, **odog2**: luminance + red-blue, **odog3**: luminance + red-blue + green-purple. Note that the number of independent matches using the decorellating filters is far higher than independent DOG filters on RGB.
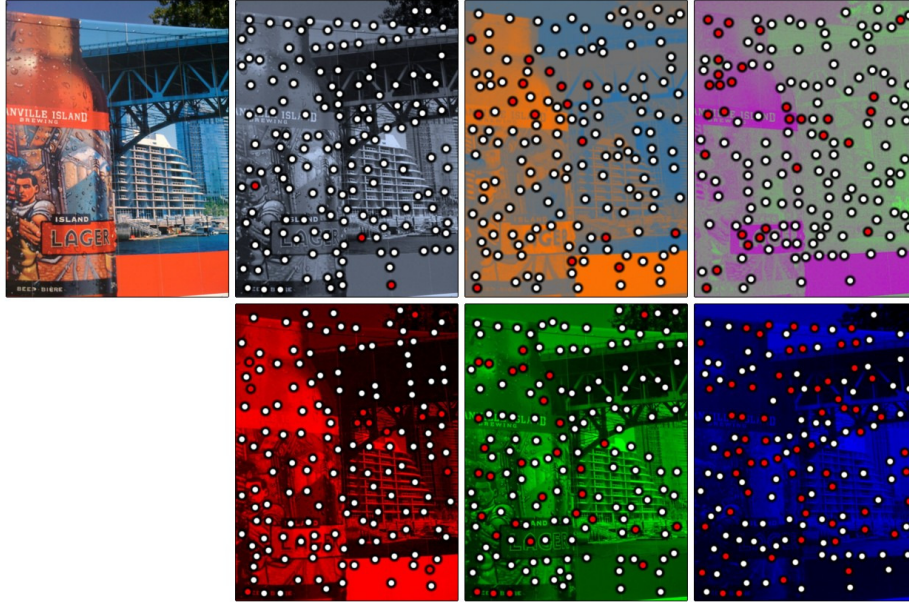
15

Figure 8: Colour Interest Points. Decorrelating the colour channels leads to a larger number of independent DOG detections (401 compared to 343 using RGB in this example). Duplicate detections are shown in red.

decorrelation. Using this formulation, we derived colour lateral inhibition filters that resemble those found in the humna visual system. We have also tested our decorrelation filters as interest point detectors, finding that the new detectors find more independent matches than difference of Gaussian filters on RGB.

5. Acknowledgements

References

[1] A. Bell and T. Sejnowski. Edges are the 'Independent components' of natural scenes. In *Advances in Neural Information Processing Systems*, pages 831–837. MIT Press, 1996.

[2] M. Brown, R. Szeliski, and S. Winder. Multi-image matching using multi-scale oriented patches. In *Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, June 2005.

[3] G. Buchsbaum and A. Gottschalk. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal Society*, B(220):89–113, 1983.

[4] M. Drew and S. Bergner. Spatio-chromatic decorrelation for color image compression. *Signal Processing: Image Communication*, 23(8):599–609, September 2008.

[5] P.-E. Forssén. Maximally stable colour regions for recognition and matching. In *Computer Vision and Pattern Recognition (CVPR'07)*, Minneapolis, USA, June 2007.

[6] A. Gooch, S. Olsen, J. Tumblin, and B. Gooch. Color2grey: Saliency-preserving color removal. In *ACM Transactions on Graphics (SIGGRAPH2005)*, volume 24.

[7] C. Harris. Geometry from visual motion. In A. Blake and A. Yuille, editors, *Active Vision*, pages 263–284. MIT Press, 1992.

[8] A. Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

[9] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[10] A. Olmos and F. Kingdom. Mcgill calibrated colour image database. http://tabby.vision.mcgill.ca, 2004.

[11] M. Ranzato, A. Krizhevsky, and G. Hinton. Factored 3-way restricted boltzmann machines for modeling natural images. In *AISTATS 2010*, Italy, 201.

[12] F. Ratliff. *Mach Bands: Quantitative Studies on Neural Networks in the Retina*. Holden-Day, San Francisco, 1965.

[13] D. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Physical Review Letters*, 73(6):814–818, August 1994.

[14] D. Ruderman, T. Cronin, and C. Chiao. Statistics of cone responses to natural images: Implications for visual coding. *Journal of the Optical Society of America*, 15(8):2036–2045, August 1998.

[15] J. Stöttinger, N. Sebe, T. Gevers, and A. Hanbury. Colour interest points for image retrieval. In *12th Computer Vision Winter Workshop*, pages 83–91, February 2007.

[16] A. Torralba and A. Oliva. Statistics of natural image categories. *Network: Computation in Neural Systems*, 14:391–412, May 2003.