

GRADIENT FIELD CORRELATION FOR KEYPOINT CORRESPONDENCE

Zeynep Engin, Melvin Lim, Anil Anthony Bharath

Imperial College London, Department of Bioengineering
South Kensington Campus, London SW7 2AZ, UK
E-mail: zeynep.engin@imperial.ac.uk

ABSTRACT

This paper presents an alternative approach to existing and widely used correlation metrics through the use of orientation information. The gradient field correlation method presented here utilises Derivative of Gaussian (DoG) operators for estimating directional derivatives of an image for two matching applications: classical planar object detection and point correspondence matching. The experimental results confirm that a suitably normalised gradient vector field, which emphasises gradient direction information in an image, leads to better selectivity when applied to classical template matching problems. For the case of establishing point correspondences, combinations of gradient vector field metrics yield higher in-lying match percentages (by RANSAC) relative to normalised cross-correlation with little extra computational cost, particularly at smaller patch sizes. It is also shown that pixel-wise field component normalisation is critical to the success of this approach.

Index Terms— Gradient vector fields, correlation, template matching, object detection, point correspondence

1. INTRODUCTION

Many matching and object recognition problems in computer vision require a measure of the quality of the match, or a match metric between an image template and the possible image patches in which the template is to be located. In the case that the instance of the template might be found anywhere in a target image, the use of correlation as the measure of match leads naturally to the use of 2D cross-correlation, which is widely used in block-based motion estimation, as well as being a generic, simple method for planar shape recognition. Its widespread use is partly because of its ease of implementation, and because Fourier theory allows cross-correlation on large patch sizes to be efficiently performed in the discrete Fourier domain.

Patch correspondence is used as the basis of recovering 3D geometry from multiple camera views [1], for super-resolution [2, 3], for image mosaicing, and may also be used for appearance-based image retrieval. Furthermore, as a match method, no prior information is required, and the technique of cross-correlation thus often serves as a component in more sophisticated techniques for feature extraction and matching.

The use of orientation information in correlation based matching methods has been suggested in the literature [4, 5, 6, 7] as an alternative to the intensity correlation methods, giving more emphasis to the ‘form’ rather than the fragile brightness constancy assumption. Orientation based matching is also used as the basis of Scale

Thanks to the Strategic Reserve Initiative of the Faculty of Engineering, Imperial College London for partial funding.

Invariant Feature Transform (SIFT) [8], a very popular technique for achieving partial invariance to scale and orientation changes between template and target image.

Our purpose in this paper is to suggest and evaluate the incorporation of orientation information in correlation matching using little more than the standard accumulate and sum primitives. This makes the algorithm implementation particularly simple: the similarity metric is extracted from field components of gradient field images using accumulate and sum operations, and with no early thresholding. The use of gradient information improves the selectivity of the algorithms as it increases the weighting given to the shape of the object or region being matched as conveyed through gradient orientations.

2. CORRELATION METRIC

The basis of using correlation as a pattern matching method lies in determining the degree to which the object under examination resembles that contained in a given reference image. The degree of resemblance is a simple statistic on which to base decisions about the object (further discussions on this can be found in [9]). The so called *normalised cross-correlation* method is a widely used match measure in correlation based pattern recognition. For digitised image patches f and g , the normalised cross-correlation measure of match is defined as

$$M(f, g) = \frac{\sum_{(i,j)} f_{ij} \cdot g_{ij}}{\sqrt{\sum_{(i,j)} f_{ij}^2} \cdot \sqrt{\sum_{(i,j)} g_{ij}^2}} \quad (1)$$

In practice, the algebraic mean of each patch is removed prior to computing this metric. Patches taken from different images are thus considered as vectors and the ‘dot product’ estimation finds the angle between these vectors in N dimensional space where N is the number of pixels in a patch. Normalisation by the standard deviations is equivalent to a normalisation by vector length, and gives the cosine of the angle between the vectors as a bounded real number between -1 and 1.

Cross correlation also has a well known statistical interpretation found in multivariate analysis, and may be shown to represent an underlying multivariate Gaussian model of statistical variability for the template and patch in which the deviation along each component (pixel) of the Gaussian is assumed to have equal variance and diagonal covariance matrices (pixel-wise independence).

3. GRADIENT FIELD MATCHING

Oriented filters, such as spatial Gabor functions, are thought to play a significant role in biological vision, and particularly for distinguishing form (shape). A generic probabilistic framework has previously

been suggested in [7] for locating planar shapes from gradient field observations. In this work, we assume a set of vector samples at known 2D image locations (\mathcal{B}, \mathcal{X}) and a Gaussian form for dispersion in the magnitude of the vector difference between the observed field and the expected gradient field, $\mu(\mathbf{x}^{(i)}|S, \mathbf{t}_c; \Theta_0)$. A plausible model for the joint likelihood function is thus derived as

$$f((\mathcal{B}, \mathcal{X})|\mathbf{t}_c) = \frac{1}{K_2} \prod_{i=1}^N \exp\{|\mathbf{b}(\mathbf{x}^{(i)}) - \mu(\mathbf{x}^{(i)}|\mathbf{t}_c)|^2/2\sigma_b^2\} \quad (2)$$

where $\mathbf{b}(\mathbf{x}^{(i)})$, $i \in \{1, \dots, N\}$, are the gradient vector field observations at pixel locations $\mathbf{x}^{(i)}$, K_2 is the normalising constant for a multivariate Gaussian, and σ_b is the standard deviation of the magnitude deviation. Equation (2) assumes that the observed field differs from some *expected field* $\mu(\mathbf{x}^{(i)}|\mathbf{t}_c)$, which in turn is dependent on the operators used to estimate image gradients and is conditioned on shape position (\mathbf{t}_c). Taking the natural logarithm of Equation (2), and discarding the small variations caused by the template and the constant terms, the main inference problem to locate a known object is given as follows:

$$\mathbf{t}_{opt} = \arg \max_{\mathbf{t}_c} \left(\sum_{i=1}^N \mathbf{b}(\mathbf{x}^{(i)}) \cdot \mu(\mathbf{x}^{(i)}|\mathbf{t}_c) \right) \quad (3)$$

In this paper, the first partial derivatives of a two dimensional Gaussian, taken along the x and y Cartesian axis directions are used for the estimation of the gradient fields. First, an observation about

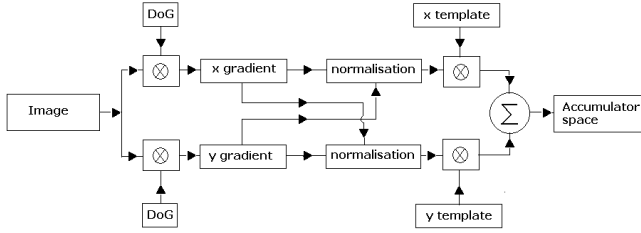


Fig. 1. Methodology for normalised gradient template matching

Equation (3): for each candidate position vector, \mathbf{t}_c , the function to be maximised is computed by summing the dot products between the gradient field values of the observed image space and the expected gradient field for the given shape around that candidate point. For scalar fields, this collapses to cross-correlation, which upon scalar fields is less selective than that of a vector field matching operation. A further improvement to this method is achieved by including a divisive normalisation term on gradient components at each pixel which is derived in [7]. A comparison of divisive normalisation with “standard” feature vector normalisation is presented in Table 1. This normalisation can be shown to correspond to a statistical model of gradient field behaviour in which object background contrast variations drawn from Gaussian distributions are permitted. Applying the normalisation to Equation (3) and considering that $\mathbf{b} = (b_x, b_y)$ and $\mu = (\mu_x, \mu_y)$, the (single-object) detection problem is then expressed as

$$\mathbf{t}_{opt} = \arg \max_{\mathbf{t}_c} \left(\sum_{i=1}^N \frac{b_x(\mathbf{x}^{(i)})\mu_x(\mathbf{x}^{(i)}|\mathbf{t}_c) + b_y(\mathbf{x}^{(i)})\mu_y(\mathbf{x}^{(i)}|\mathbf{t}_c)}{Z_0^{(i)}} \right) \quad (4)$$

and the normalisation term in the denominator is defined as

$$Z_0^{(i)} = \alpha + |\mathbf{b}(\mathbf{x}^{(i)})| = \alpha + \sqrt{b_x^2(\mathbf{x}^{(i)}) + b_y^2(\mathbf{x}^{(i)})} \quad (5)$$

where α is a constant factor and is proportional to the highest intensity value in the input images. Note that the template (or the expected gradient field) $\mu(\mathbf{x}^{(i)}|\mathbf{t}_c)$ is assumed to be previously normalised. It is emphasised that the normalisation here is quite different to the denominator of Equation (1) because the components of the gradient vector at *each* pixel are individually normalised to have unit length before the summation is performed across the patch.

4. EXPERIMENTS

4.1. Object detection experiments

The first practical performance issue we address in natural images is the selectivity of matching in vector fields, and the noise stability under quite typical conditions, on a rather difficult matching problem of matching a *specific* pen cover within a cluttered image (see Fig. 2). The experiments start with converting the original colour

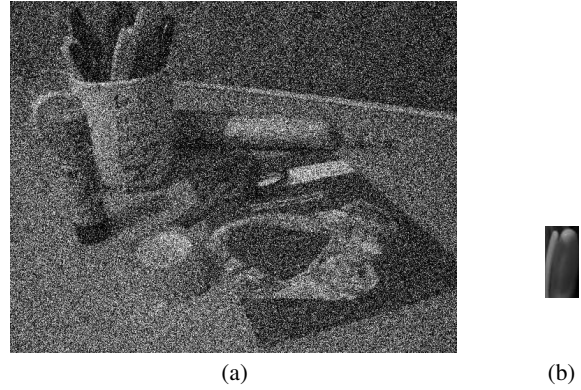


Fig. 2. The detection problem, (a) Cluttered input image (b) Template (pencover) taken from a separate image, therefore contains some distortion

image into a grey-scale representation, in this case, by extracting the green channel. For this matching problem where one of the pen covers is sought, the mask is taken from a different image under a small degree of affine transformation. The studies are carried over images with four different Gaussian noise levels, and for each noise level fifty experiments are performed. For each noisy image, three methods are applied: scalar cross-correlation, gradient vector matching and gradient magnitude matching (Compact Hough Transform with magnitude weighting, without direction). Cross-correlation is applied using a mean subtracted template image. For the gradient vector matching, the gradient components are estimated by an 11×11 DoG operator for both the image and the mask. The accumulator space is constructed by using the vector matching method of Equation (3), the procedure is illustrated in Fig. 1. For both methods of correlation and gradient vector matching, the absolute values of the accumulator spaces are used during the comparison. Finally, the magnitude correlation is performed after estimating the magnitudes of the gradient vectors using the same DoG operators as for the gradient vector matching. The comparison is done over the principal global maximum and the next highest (false) local maximum in the accumulator spaces. The ratios of measurements of the false maxima to the true maxima without normalisation are illustrated in Fig. 3(a) and Fig. 3(b) and for simplicity¹. The error bars show the minimum

¹Note that a ratio is favoured because the resulting numbers are easy to interpret. From a probabilistic perspective, a subtraction is, strictly speaking,

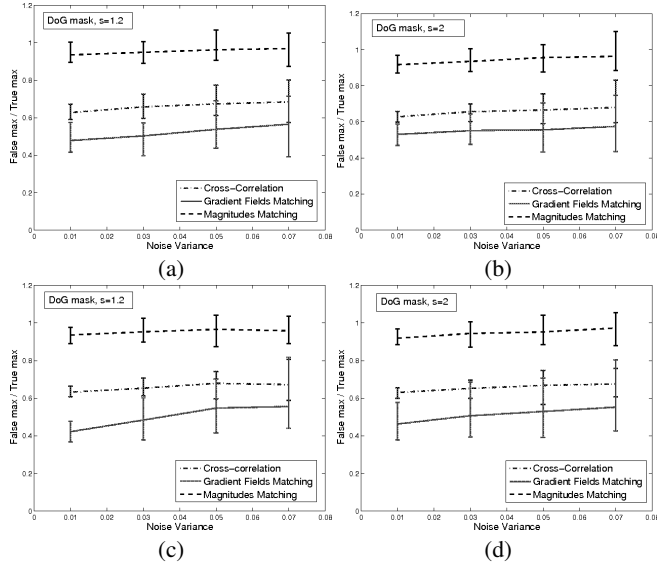


Fig. 3. (a) Inverse Ratio of true maxima to next highest false maxima for 50 runs at each of four different noise levels, using a value of σ (scale parameter) in the Derivative of Gaussian (DoG) of 1.2 pixels. (b) Inverse ratio of true maxima to next highest false maxima for 50 runs at each of four different noise levels, using a value of σ of 2 pixels. (c) Inverse Ratio of true maxima to next highest false maxima with the normalisation for 50 runs at each of four different noise levels, using a value of σ (scale parameter) in the DoG of 1.2 pixels. (d) Inverse Ratio of true maxima to next highest false maxima with the normalisation for 50 runs at each of four different noise levels, using a value of σ (scale parameter) in the DoG of 2 pixels.

and maximum deviations of the match metric space ratios across the fifty runs at one noise level. It is observed that although stability in noise might be slightly higher for correlation, the shape selectivity is higher for the matching performed on vector fields. Indeed, one surprising result is that gradient *magnitude matching* is very sensitive to noise arising from large amplitude gradient values that contain inconsistent boundary direction information; this suggests strongly against the use of non-directional Compact Hough Transform based methods for shape matching. Stability in noise improves with the use of a larger blurring parameter on the gradient estimators (note narrower error bars on the gradient vector matching curve in Fig. 3(a) compared to Fig. 3(b)), but this comes at the cost of loss in selectivity, shown by the curves for Correlation and Gradient vector matching getting closer together in Fig. 3(b). This is improved when the normalisation term described by Equation (4) and Equation (5) is included. Fig. 3(c) and Fig. 3(d) show the performance of the normalised algorithm. Inclusion of the normalisation term results in overall improvement but it becomes more significant as the blurring increases.

4.2. Key point correspondence experiments

For correspondence, when two different views of the same scene are given, the problem is to match each location in the first image with the correct corresponding location in the second one. In computer vision, it is common to estimate the parameters of a geometric

more appropriate, as this space is a logarithmic one.

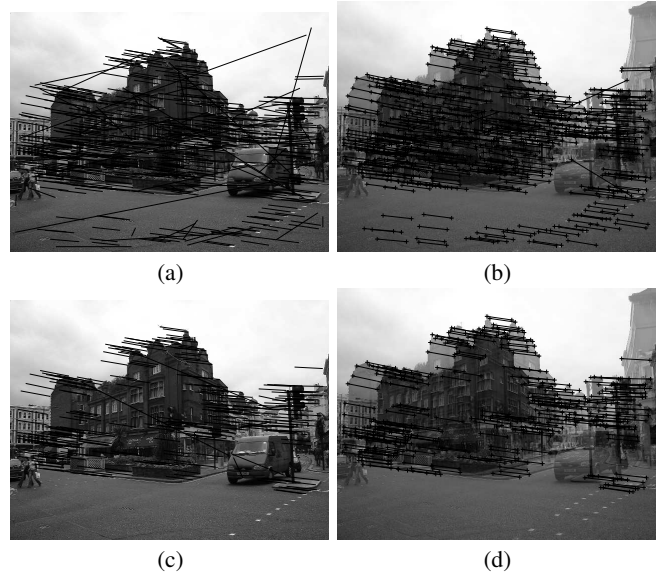


Fig. 4. (a) Putative matches obtained using only cross-correlation as the similarity metric (567 matches). (b) Inliers of correlation based matching after running the RANSAC algorithm (512 matches). (c) The putative matches (282 matches) running the orientation matching algorithm (DoG mask $\sigma = 2$). (d) Inliers of orientation based matching after running the RANSAC algorithm (274 matches).

Table 1. Normalisation Comparison

	Divisive normalisation			Feature vector normalisation		
	Put	Inl	Ratio	Put	Inl	Ratio
South Kensington	304	279	0.9178	612	518	0.8464
Drawer	245	207	0.8449	499	364	0.7295
Desk	172	163	0.9477	229	213	0.9301
St. Pancras	287	263	0.9164	602	498	0.8272

transformation, such as a homography, by automatic detection and to analyse corresponding features among the input images. The robustness and speed of this estimation is heavily dependent on the quality of the initial determination of inlying matches in the different images examined. Here, following the methodology described by Hartley and Zisserman [1], in each image several hundred “interest points” are automatically detected with subpixel accuracy using Kovese’s implementation of the Harris feature detector [10, 11]. Putative correspondence locations are identified by comparing the image neighbourhoods around the features, using a similarity metric. These correspondences are refined using a Random sample consensus (RANSAC) algorithm [12] that extracts “inliers” whose interimage motion is consistent with a homography.

The choice of a metric for initial putative matches which uses not only the intensity information, but also the orientation information in the keypoint’s neighbourhood will give a more robust description of these points. Different similarity metrics (obtained by modifying the correlation matrix C) used for the experiments in this section are those of normalised cross-correlation ($C = C_1$), gradient field correlation using two different DoG scales ($C = C_{g1}$ or $C = C_{g2}$) and a combination of these gradient fields correlation ($C = C_{g1} + C_{g2}$).

Fig. 4 illustrates the quality of the match for the cases of correlation and the gradient vector matching. Fig. 4(a) shows the putative matches for the image pairs when cross-correlation is used as

Table 2. Correspondence Performance

	patch s = 7		patch s = 11	
	Drawer	South Ken.	Drawer	South Ken.
Correlation	0.5946	0.7128	0.7261	0.7997
Gra. ($\sigma_1 = 1.2$)	0.7819	0.8724	0.8396	0.9171
Gra. ($\sigma_2 = 3$)	0.6312	0.7833	0.6500	0.9123
Comb.Gra.	0.7528	0.9086	0.8528	0.9509
	patch s = 19		patch s = 23	
	Drawer	South Ken.	Drawer	South Ken.
Correlation	0.7798	0.9040	0.7644	0.9161
Gra. ($\sigma_1 = 1.2$)	0.8708	0.9398	0.8511	0.9357
Gra. ($\sigma_2 = 3$)	0.7918	0.9467	0.8597	0.9530
Comb.Gra.	0.8779	0.9621	0.8883	0.9646
	patch s = 35		patch s = 45	
	Drawer	South Ken.	Drawer	South Ken.
Correlation	0.8283	0.9473	0.8621	0.9547
Gra. ($\sigma_1 = 1.2$)	0.8842	0.9553	0.8771	0.9548
Gra. ($\sigma_2 = 3$)	0.8893	0.9638	0.8988	0.9849
Comb.Gra.	0.8879	0.9645	0.8804	0.9725

the similarity metric for a patch size of 19×19 pixels around each feature point and Fig. 4(b) shows the inliers after running Kovese's [11] RANSAC implementation. The same operations are applied for Fig. 4(c) and Fig. 4(d) with the difference being the use of gradient information in the similarity metric. The gradient fields here are obtained using a DoG function at scale 2. Table 1 shows the performance of two different normalisation alternatives in the correspondence algorithm for different input image pairs². It is evident from the table that the use of 'divisive normalisation' as described by Equation (5) yields better results than 'feature vector normalisation' which might appear as corresponding to the normalisation in the original cross-correlation measure given by Equation (1) for the gradient vector fields case. An evaluation of the effect for different sizes of DoG operators (scales of 1.2 and 3) and patch sizes (between 7 and 45 pixels) is given for two different images in Table 2. Experiments have been repeated ten times for each case, and the ratio of the number of inliers to the number of putative matches was used as a measure of performance in each method. Loosely, it can be seen from the table that the individual channels obtained from gradient fields consistently outperform the intensity cross-correlation methods. The combination of two scales of gradient estimation also gives rise to improved putative match selectivity at smaller patch sizes.

5. CONCLUSION

A gradient field matching method has been presented as an alternative to the typical implementation of intensity correlation based matching. The method consists of very low-complexity operations such as accumulation and multiplication, with a simple non-linearity that achieves pixel-wise normalisation of gradient field components. Although as the correlation patch size is increased, the relative advantage of the gradient field approach does reduce, it is clear that including the outputs of even simple gradient estimators can dramatically improve the correspondence quality at small patch sizes.

The techniques presented here do not address the problems of rotation and scale invariance, or of determining correspondence in the case of general affine transformations. Yet, the low complexity and robustness of the approach presented here still makes it appropriate for many problems of object recognition and keypoint correspondence, such as in industrial inspection and some of the less challenging geometric recovery problems. Furthermore, employing a variety of scaled or rotated templates, a technique used to extend both

correlation-based template matching and the Compact Hough Transform to handle rotations and scalings of a sought shape, can also be applied to address the scale and rotation problem in the vector-field gradient matching scheme.

Further work will involve more extensive evaluations across a wide range of images, and will consider the extension of the principle using multiple directional subbands in a wavelet framework, such as either the steerable complex wavelet scheme of Bharath [13], or the very efficient modified DTCWT of Kingsbury [14].

6. REFERENCES

- [1] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.
- [2] D. Capel and A. Zisserman, "Computer vision applied to super resolution," *Signal Processing Magazine, IEEE*, vol. 20, no. 3, pp. 75–86, 2003.
- [3] C. P. Sung, K. P. Min, and G. K. Moon, "Super-resolution image reconstruction: a technical overview," *Signal Processing Magazine, IEEE*, vol. 20, no. 3, pp. 21–36, 2003.
- [4] A. Fitch, A. Kadyrov, W. Christmas, and J. Kittler, "Orientation correlation," in *British Machine Vision Conference*, P. Rosin and D. Marshall, Eds., vol. 1, 2002, pp. 133–142. [Online]. Available: citeseer.ist.psu.edu/540764.html
- [5] F. Ullah, S. Kaneko, and S. Igarashi, "Orientation code matching for robust object search," *IEICE Transactions on Information and Systems*, no. 8, pp. 999–1006, 2001.
- [6] D. Scharstein, "Matching images by comparing their gradient fields," in *ICPR94*, 1994, pp. A:572–575. [Online]. Available: citeseer.ist.psu.edu/scharstein94matching.html
- [7] S. Basalamah, A. Bharath, and D. McRobbie, "Contrast marginalised gradient template matching," *Lecture Notes In Computer Science*, vol. 3023, pp. 417–429, 2004.
- [8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, NOV 2004.
- [9] B. Kumar, A. Mahalanobis, and R. D. Juday, *Correlation Pattern Recognition*. Cambridge University Press, ISBN: 0521571030, 2005.
- [10] C. Harris and M. Stephens, "A combined corner and edge detector," *Proc. Alvey Vision Conf.*, pp. 147–151, 1988.
- [11] P. Kovese, "RANSAC algorithm." [Online]. Available: <http://www.csse.uwa.edu.au/pk/research/matlabfns/>
- [12] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Comm. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [13] A. A. Bharath and J. Ng, "A steerable complex wavelet construction and its application to image denoising," *IEEE Transactions on Image Processing*, vol. 14, no. 7, pp. 948–959, July 2005.
- [14] N. Kingsbury, "Rotation-invariant object recognition using edge-profile clusters," in *European Conference on Signal Processing (EUSIPCO)*, 2006.

²Images at "<http://www.bg.ic.ac.uk/research/vision/CorrespImgs.html>"