Invariant Template Matching with Tangent Vectors

Brendt Wohlberg

T-7, MS-B284
Los Alamos National Laboratory
Los Alamos, NM 87545
Tel: (505) 667 6886
Fax: (505) 665 5757
Email: `brendt@t7.lanl.gov`

Kevin Vixie

T-7, MS-B284
Los Alamos National Laboratory
Los Alamos
NM 87545
Tel: (505) 665 9887
Fax: (505) 665 5757
Email: `vixie@lanl.gov`

Abstract. *Template matching* is the search for a known object, represented by a *template* image, at an arbitrary location within a larger image. The local measure of match is often desired to be invariant to certain transforms, such as rotation and dilation, of the template. While a variety of solutions have been proposed, most are designed to provide invariance to a specific transform or set of transforms, and often involve significant computational demands. When invariance to "small" transformations of the template (e.g. rotation by a small angle) is sufficient, local linear approximations to these transforms may be used to allow template matching with invariance to arbitrary transforms, without significantly increased computational requirements. *DOI: 10.0000/XXXX*

Subject terms: Template matching, correlation, invariance, linear approximation

Manuscript compiled October 10, 2006

1 Introduction

Template matching is the location of a known object within an image by computing a *measure of match* between a *template* representing the object to be located, and all template-sized subregions of the image of interest [1, pp. 400-404] [2, pp. 651-653]. The measure of match is computed at every point in the image to construct the match surface, as indicated in Figure 1. Template matching is related to *matched filters*, which are optimal for detecting a deterministic signal with additive white noise, when the measure of match is correlation based.

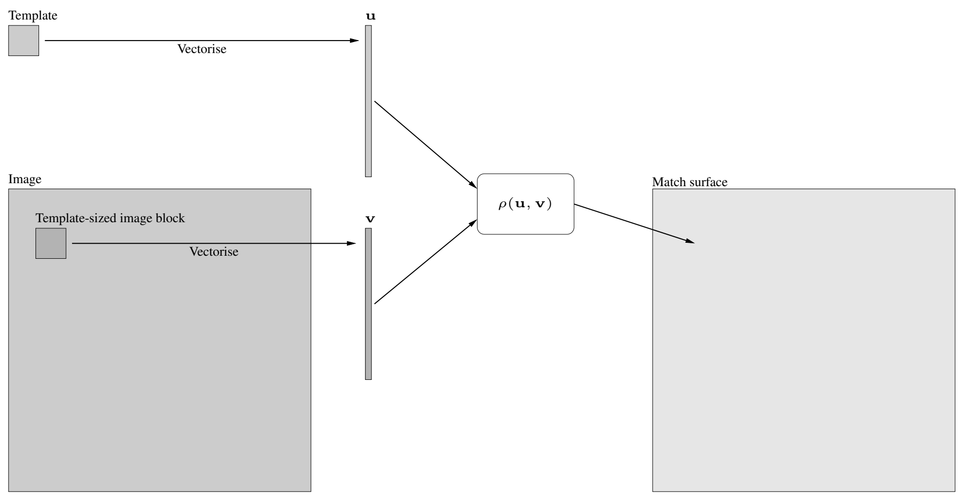


Fig. 1 Computing the match surface for measure of match $\rho(\mathbf{u}, \mathbf{v})$.

Consider vectors $\mathbf{u}$ and $\mathbf{v}$ representing a template and an image block respectively (the vector representation being constructed, for example, by a raster-scan of the image block), and define the mean-subtracting projection

$$R\mathbf{u} = \mathbf{u} - \frac{\langle \mathbf{u}, \mathbf{1} \rangle}{\langle \mathbf{1}, \mathbf{1} \rangle} \mathbf{1},$$

where $\mathbf{1}$ is a vector of unit components, so that the correlation coefficient (the cosine of the angle between $\mathbf{u}$ and $\mathbf{v}$ after subtraction of their means) may be written as

$$\rho_{\mathrm{cor}}(\mathbf{u}, \mathbf{v}) = \frac{\langle R\mathbf{u}, R\mathbf{v} \rangle}{\|R\mathbf{u}\| \, \|R\mathbf{v}\|}.$$

This widely used measure of match is invariant to scaling and level shifts of pixel values.

Direct computation of the correlation surface (the match surface with a correlation-based measure of match) is computationally expensive. An efficient implementation can be derived by noting that the inner product between a template and each position in an image can be expressed as the convolution of the vertically and horizontally flipped template and the image. Once expressed as a convolution, the DFT convolution theorem allows the use of FFTs to implement the convolutions efficiently.[3]

In most applications, one would like to detect occurrences of the template subject to certain transforms, in particular, rotation and scaling. A wide variety of techniques for rotation and scale invariant template matching have been considered, the two main types being based on the Fourier-Mellin transform[4] and on invariant moments.[5] Most of these methods are designed for invariance with respect to a specific small set of transforms, and many are significantly more computationally expensive than simple correlation-based template matching.

An alternative strategy, which leads to a very general method suitable for arbitrary transforms, is motivated by considering the manifold, in the vector space of template images, generated by the application of the transform(s) of interest to the primary template. Invariant matching may then be achieved, in principle, by determining whether an image block lies within this manifold. Since the difficulty of appropriately representing a general manifold makes direct application of this idea impractical, it is necessary to approximate the manifold by a linear subspace. Given this approximation, one may easily determine whether an image block lies within the subspace by projecting it into that subspace and comparing with the original image block. Computational efficiency is achieved by constructing an orthonormal basis of templates for the linear subspace; the projection and measure of match may then be computed in terms of the correlation between the image and each of the basis templates, so that the computation expense is of the order of the number of basis templates greater than the expense of matching with a single template. Uenohara and Kanade[6] have described such an approach in some detail, using the Karhunen-Loève Transform (KLT) to construct the orthonormal basis. (The basic idea of the KLT approach has also been considered in earlier work,[7] and has been shown to be optimal in a matched filtering sense for a stochastic ensemble generated by template distorting transforms.[8,9]) We shall denote the correlation computed using this approach by $\rho_{\mathrm{klt}}(\cdot, \cdot)$.

2 Tangent Approximations to Invariances

While the general approach of Uenohara and Kanade[6] may be applied for arbitrary transforms, they only describe the case of rotation, constructing their template set by one degree rotations of the primary template, and do not address the problem of the potential explosion of the template set size when using multiple transforms. Instead of the KLT method for constructing the linear subspace for matching, we propose a method inspired by the *tangent distance* classification technique of of Simard *et al.*.[10] While inspired by this work (but more naturally presented as an alternative to the work of Uenohara and Kanade,[6] of which we became aware at a later stage), we emphasise that our approach is not a trivial extension thereof; we use only the idea of tangent vectors to transforms, and not the specific machinery for computing distances. Direct application of the tangent distance is inappropriate for template matching since it is unnormalised, and is not invariant to scaling and level shifts of pixel values. In addition, the solution of two sets of linear systems[10] for every point in the match surface represents an undesirable computational overhead.

Instead of explicitly building a potentially large set of transformed templates, we restrict our method to account for "small" transformations, so that a local linear approximation to these transforms justifies using the tangent vectors to the transforms to construct the representative subspace. First consider invariance with respect to a function $f(\mathbf{u}, \alpha)$, where $\mathbf{u}$ is the (vectorised) image block to which the transform is applied, and α is the single transform parameter (for example, angle of rotation if f is a rotation transform). We assume that $\alpha = 0$ gives the identity transform, so that $f(\mathbf{u}, 0) = \mathbf{u}$. Expanding f in a Taylor series about $\alpha = 0$, we obtain

$$f(\mathbf{u}, \alpha) = f(\mathbf{u}, 0) + \left.\frac{\partial f}{\partial \alpha}\right|_{\alpha=0} \alpha + \left.\frac{\partial^2 f}{\partial \alpha^2}\right|_{\alpha=0} \frac{\alpha^2}{2} + \dots \quad .$$

Defining

$$\mathbf{t}(\mathbf{u}) = \left.\frac{\partial f(\mathbf{u}, \alpha)}{\partial \alpha}\right|_{\alpha=0},$$

we may then write $f(\mathbf{u}, \alpha) \approx \mathbf{u} + \alpha \mathbf{t}(\mathbf{u})$ for sufficiently small $|\alpha|$. In general, we consider a transformation $f(\mathbf{u}, \alpha_0, \alpha_1, \dots, \alpha_{L-1})$ with multiple parameters, retaining the assumption that $f(\mathbf{u}, 0, 0, \dots, 0) = \mathbf{u}$. In this case we define

$$\mathbf{t}_k(\mathbf{u}) = \left.\frac{\partial f(\mathbf{u}, \alpha_0, \alpha_1, \dots \alpha_{L-1})}{\partial \alpha_k}\right|_{\alpha=0} \quad k \in \{0, 1, \dots, L-1\},$$

so that

$$f(\mathbf{u}, \alpha_0, \alpha_1, \dots \alpha_{L-1}) \approx \mathbf{u} + \sum_{k=0}^{L-1} \alpha_k \mathbf{t}_k(\mathbf{u})$$

for sufficiently small $|\alpha_k|$.

The linear combinations of $\mathbf{u}$ (we allow for scaling of the primary template) and the tangent vectors $\mathbf{t}_k(\mathbf{u})$ define a subspace within which small transformations of $\mathbf{u}$ may be approximated.[10] Projection of an image block into this subspace therefore provides an indication of how well the block may be approximated by a small transform of the template block. In particular, a template matching measure of match that is invariant to small transforms can be defined as the correlation between the projected and original image blocks. If the image block is indeed a small transform of the template, the projection is an identity transform, and the correlation is unity; conversely, if the correlation is significantly smaller than unity, the image block is poorly approximated in this subspace, and is therefore not similar to a small transform of the template.

The projection into the subspace of template and tangent vectors is constructed using an orthonormal basis for this subspace. First, we construct the tangent vector set $\{\mathbf{t}_k(R\mathbf{u})\}_{k \in \{0,1,\dots,L-1\}}$. Defining the projection orthogonal to both $R\mathbf{u}$ and $\mathbf{1}$ as

$$Q_{\mathbf{u}}\mathbf{v} = R\mathbf{v} - \frac{\langle R\mathbf{v}, R\mathbf{u} \rangle}{\langle R\mathbf{u}, R\mathbf{u} \rangle} R\mathbf{u},$$

we orthogonalise and normalise the set $\{Q_{\mathbf{u}}\mathbf{t}_k(R\mathbf{u})\}_{k\in\{0,1,\ldots,L-1\}}$, labelling the resulting set $\{\tilde{\mathbf{t}}_k\}_{k\in\{0,1,\ldots,L-1\}}$, and define $\tilde{\mathbf{u}} = \frac{R\mathbf{u}}{\|R\mathbf{u}\|}$, providing us with the basis $\{\tilde{\mathbf{u}}, \tilde{\mathbf{t}}_k\}_{k\in\{0,1,\ldots,L-1\}}$ for the subspace of small transformations of $R\mathbf{u}$. The projection of any image block $\mathbf{v}$ into this subspace is therefore

$$P_{\mathbf{u},f}\mathbf{v} = \langle \mathbf{v}, \tilde{\mathbf{u}}\rangle \tilde{\mathbf{u}} + \sum_{k=0}^{L-1} \langle \mathbf{v}, \tilde{\mathbf{t}}_k\rangle \tilde{\mathbf{t}}_k.$$

Using this notation, the invariant measure of match (described above) between image block $\mathbf{v}$ and template $\mathbf{u}$ can be written $\rho_{\text{cor}}(\mathbf{v}, P_{\mathbf{u},f}\mathbf{v})$. Using the following equalities ($\mathbf{v}$ is an arbitrary vector)

$$\begin{array}{lll} \langle R\mathbf{v}, \mathbf{1}\rangle = \langle P_{\mathbf{u},f}\mathbf{v}, \mathbf{1}\rangle = 0 & \langle \tilde{\mathbf{u}}, \tilde{\mathbf{t}}_k\rangle = 0,\ \|\tilde{\mathbf{u}}\| = \|\tilde{\mathbf{t}}_k\| = 1\ \ \forall k \in \{0, 1, \ldots, L-1\} & \\ RP_{\mathbf{u},f}\mathbf{v} = P_{\mathbf{u},f}\mathbf{v} & \langle R\mathbf{v}, P_{\mathbf{u},f}\mathbf{v}\rangle = \langle \mathbf{v}, P_{\mathbf{u},f}\mathbf{v}\rangle & \|P_{\mathbf{u},f}\mathbf{v}\|^2 = \langle \mathbf{v}, P_{\mathbf{u},f}\mathbf{v}\rangle = \langle \mathbf{v}, \tilde{\mathbf{u}}\rangle^2 + \sum_{k=0}^{L-1}\langle \mathbf{v}, \tilde{\mathbf{t}}_k\rangle^2, \end{array}$$

one may easily derive that

$$\rho_{\text{cor}}(\mathbf{v}, P_{\mathbf{u},f}\mathbf{v}) = \frac{\|P_{\mathbf{u},f}\mathbf{v}\|}{\|R\mathbf{v}\|},$$

which may be interpreted as the fraction of the energy of $R\mathbf{v}$ remaining after its projection into the subspace spanned by the template and its corresponding tangent vectors.

Since this measure consists of a ratio of (positive) vector magnitudes, it does not distinguish between positive and negative correlations. This deficiency is easily remedied by multiplying by the sign of the inner product of $\mathbf{v}$ and $\tilde{\mathbf{u}}$, leading to the definition

$$\rho_0(\mathbf{u}, \mathbf{v}) = \text{sgn}(\langle \mathbf{v}, \tilde{\mathbf{u}}\rangle)\frac{\|P_{\mathbf{u},f}\mathbf{v}\|}{\|R\mathbf{v}\|} = \text{sgn}(\langle \mathbf{v}, \tilde{\mathbf{u}}\rangle)\sqrt{\frac{\langle \mathbf{v}, \tilde{\mathbf{u}}\rangle^2 + \sum_{k=0}^{L-1}\langle \mathbf{v}, \tilde{\mathbf{t}}_k\rangle^2}{\langle \mathbf{v}, \mathbf{v}\rangle - \frac{\langle \mathbf{v}, \mathbf{1}\rangle^2}{\langle \mathbf{1}, \mathbf{1}\rangle}}},$$

where

$$\text{sgn}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 1 & \text{otherwise.} \end{cases}$$

The correlation $\rho_0(\mathbf{u}, \mathbf{v})$ provides an increased response when the image block $\mathbf{v}$ is similar to the result of the specified small transformation f of the template $\mathbf{u}$. It is clear, however, that it also gives a significant response when the image block $\mathbf{v}$ is similar to one of the tangent vectors $\tilde{\mathbf{t}}_k$ of transformation f. This shortcoming is addressed by defining

$$\rho_1(\mathbf{u}, \mathbf{v}) = \frac{|\langle \mathbf{v}, \tilde{\mathbf{u}}\rangle|}{\|P_{\mathbf{u},f}\mathbf{v}\|} = \frac{|\langle \mathbf{v}, \tilde{\mathbf{u}}\rangle|}{\sqrt{\langle \mathbf{v}, \tilde{\mathbf{u}}\rangle^2 + \sum_{k=0}^{L-1}\langle \mathbf{v}, \tilde{\mathbf{t}}_k\rangle}}$$

as a measure of the fraction of the energy of $P_{\mathbf{u},f}\mathbf{v}$ in the direction of the primary template. When $\rho_0(\mathbf{u}, \mathbf{v})$ is large, a large value of $\rho_1(\mathbf{u}, \mathbf{v})$ ensures that the match indeed represents a small transform of the primary template, and not one of the tangent vectors. Instead of a two-stage approach, checking $\rho_1(\mathbf{u},\mathbf{v})$ where $\rho_0(\mathbf{u},\mathbf{v})$ is large, we construct a hybrid measure of match based on both $\rho_0(\mathbf{u},\mathbf{v})$ and $\rho_1(\mathbf{u},\mathbf{v})$. Since $\rho_0(\mathbf{u},\mathbf{v})\rho_1(\mathbf{u},\mathbf{v}) = \rho_{\text{cor}}(\mathbf{u},\mathbf{v})$ (see Appendix A) we define

$$\rho_{\text{tan}}(\mathbf{u},\mathbf{v}) = \rho_0(\mathbf{u},\mathbf{v})\eta(\rho_1(\mathbf{u},\mathbf{v})),$$

where

$$\eta(x) = \begin{cases} x & \text{if } \ x < t_0 \\ \left(\frac{1-t_0}{t_1-t_0}\right) x + \left(\frac{t_1-1}{t_1-t_0}\right) t_0 & \text{if } \ t_0 \leq x \leq t_1 \\ 1 & \text{if } \ t_1 < x \end{cases},$$

so that $\rho_{\text{tan}}(\mathbf{u},\mathbf{v})$ is the standard $\rho_{\text{cor}}(\mathbf{u},\mathbf{v})$ when $\rho_1(\mathbf{u},\mathbf{v}) < t_0$, becomes $\rho_0(\mathbf{u},\mathbf{v})$ when $\rho_1(\mathbf{u},\mathbf{v}) > t_1$, and is interpolated between the two values when $t_0 \leq \rho_1(\mathbf{u},\mathbf{v}) \leq t_1$. In the experiments described below, we set $t_0 = 0.25$ and $t_1 = 0.50$.

As discussed in Appendix B, construction of tangent vectors $\mathbf{t}_k$ requires computing derivatives of the templates. These derivatives are calculated by convolution of the template $\mathbf{u}$ with the derivatives of a Gaussian kernel, which is equivalent to taking the derivatives of $\mathbf{u}$ after smoothing by convolution with the same Gaussian kernel. This smoothing is essential to obtain stable estimates of the derivatives since images containing sharp edges are, in principle, non-differentiable. Since the tangent vectors are derived from a smoothed template, it is reasonable, for consistency, to apply the same smoothing to the primary template, and similarly, to the image to be matched. In the results presented here, we distinguish between those for unsmoothed correlations, in which the primary template and image to be matched are not smoothed, but a smoothed version of the primary template is used in computing the tangents, and smoothed correlations, in which smoothing is applied to the primary template and image to be matched, in addition to the usual smoothing applied in the tangent computation. For the ρ_{tan} correlation these unsmoothed and smoothed correlations correspond respectively to the inconsistent (tangents correspond to a smoothed version of the template) and consistent (tangents correspond to the actual template used) cases, but for correlations ρ_{cor} and ρ_{klt} there is no such interpretation, since they do not require the use of tangent vectors computed using smoothing.

3 Efficient Implementation

When used for template matching, the measures of match defined above need to be computed for every template-sized block within the image to be matched. The most expensive component of these computations is the evaluation of the inner products at every point in this image. The critical observation leading to efficient implementation is that the inner product of a template with every image block may be expressed as the convolution of the vertically and horizontally flipped template with the whole image. The DFT convolution theorem allows this to be computed, using the FFT, as the inverse DFT of the pointwise products of the DFTs of template and image respectively.[3]

Assuming an $N \times N$ image, and an $M \times M$ template, with $M \ll N$ so that the cost of the FFT on the template is negligible, the resulting algorithm for computing the inner product at every point has an O($N^2 \log_2 N$) cost.

Now, noting that $\langle R\mathbf{u}, \mathbf{1} \rangle = 0$, so that $\langle R\mathbf{u}, R\mathbf{v} \rangle = \langle R\mathbf{u}, \mathbf{v} \rangle$, and expanding out $\|R\mathbf{v}\| = \sqrt{\langle R\mathbf{u}, R\mathbf{u} \rangle}$, we have

$$\rho_{\mathrm{cor}}(\mathbf{u}, \mathbf{v}) = \frac{\left\langle \frac{R\mathbf{u}}{\|R\mathbf{u}\|}, \mathbf{v} \right\rangle}{\sqrt{\langle \mathbf{v}, \mathbf{v} \rangle - \frac{\langle \mathbf{v}, \mathbf{1} \rangle^2}{M^2}}},$$

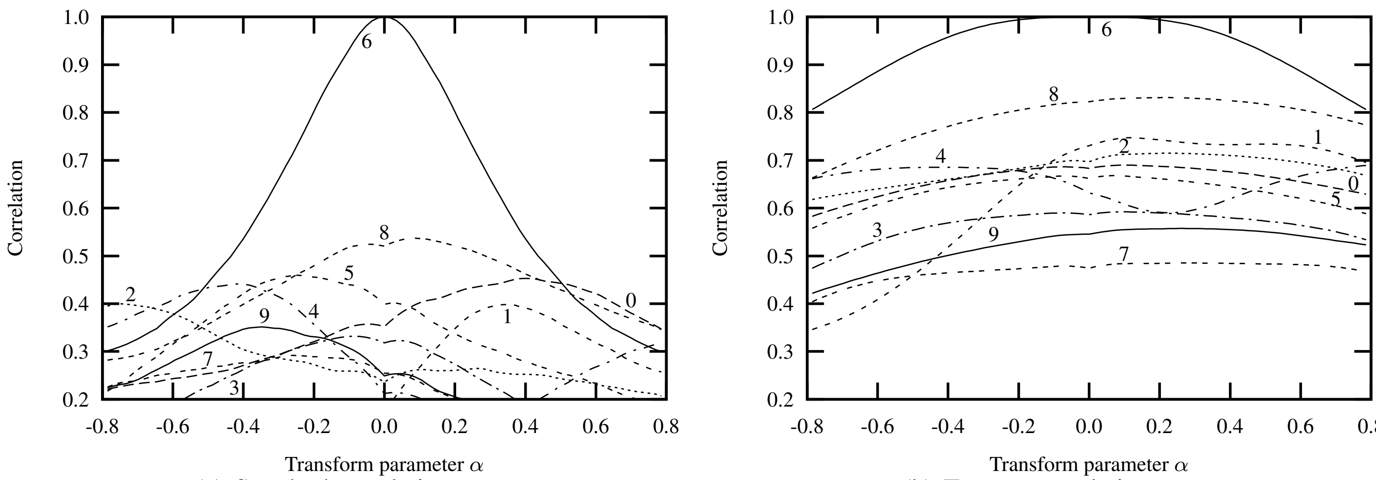
the cost of computing which is dominated by the three inner products at every point in the image. The FFT approach outlined above may be used to compute $\langle \mathbf{v}, \mathbf{1} \rangle$ for every image block $\mathbf{v}$ by convolution of the image with an image block with all unit entries. A similar approach allows computation of $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ by noting that $\langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{v} \odot \mathbf{v}, \mathbf{1} \rangle$, where $\mathbf{v} \odot \mathbf{v}$ represents the Hadamard product of $\mathbf{v}$ with itself. (These techniques are also utilised by Uenohara and Kanade.[6]) Consequently the match surface using $\rho_{\mathrm{cor}}(\cdot, \cdot)$ may be computed at an O($N^2 \log_2 N$) cost. The correlation $\rho_{\tan}(\cdot, \cdot)$ involves an additional L (the number of tangent vectors) inner products, so the corresponding match surface has an O($LN^2 \log_2 N$) cost.

4 Results

We test the proposed method using the MNIST[11,12] dataset of handwritten digits, and consider invariance to the dilation, rotation, and the diagonal hyperbolic, parallel hyperbolic, and thickening transforms which have already been demonstrated to be relevant to this type of data[10] (although not in the context of template matching). Since template matching is inherently translation invariant, the tangents to the translation transform are excluded in these experiments. We note, however, that the correlation with the translation tangent vectors can be used to provide an estimate of template location to subpixel accuracy, which may be useful in certain applications.

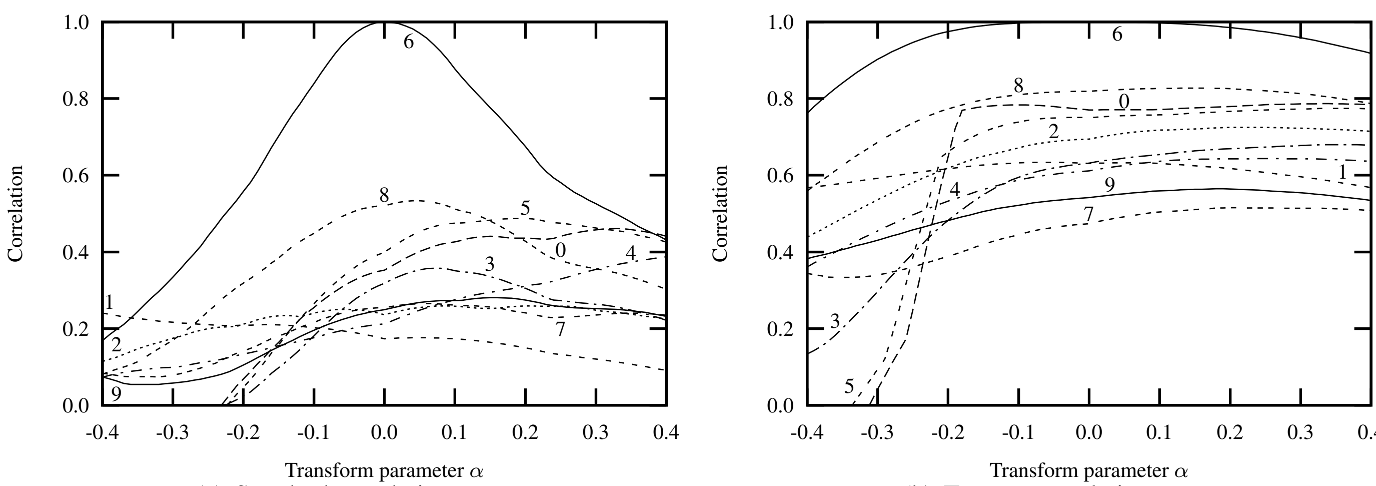
Fig. 2 A collage of the templates selected for digits 0 to 9.

Computation of the tangents for these transforms via convolution by the derivatives of a 2D Gaussian is described in Appendix B. In all experiments we use a 12×12 Gaussian kernel with standard deviation $\sigma = 1.75$, which was found experimentally to provide a good compromise between invariance (improving with increased smoothing) and feature preservation (deteriorating with increased smoothing). A representative template was selected (see Figure 2) for each of the 10 digits. Since we do not expect that the transform invariance accounts for all of the variation between examples of a specific digit, this would not be adequate for a practical classification technique, corresponding to using a *single* training example for each digit, but is appropriate for the evaluation reported here.



(a) Standard correlation ρ_{cor}. (b) Tangent correlation $\rho_{\tan}$.

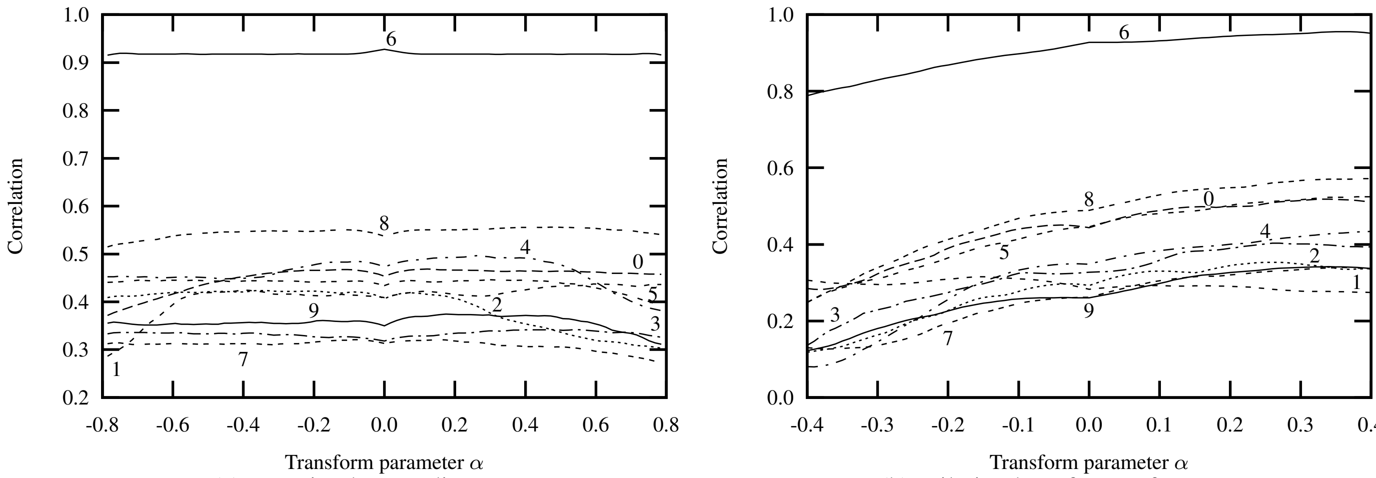
Fig. 3 Rotation transform: correlations between template (digit '6') and digit images rotated by α radians.



(a) Standard correlation ρ_{cor}. (b) Tangent correlation $\rho_{\tan}$.

Fig. 4 Dilation transform: correlations between template (digit '6') and digit images dilated by a factor of $\alpha - 1$.

In the tangent distance approach, a non-linear transform is approximated by its local tangents, which are used to modify the Euclidean distance between two vectors.[10] In our application, the tangents are also used to approximate non-linear transforms, but as a modification to a correlation rather than to a Euclidean distance. In the spirit of Figure 5 in,[10] we illustrate the benefits of the tangent-modified correlation (for rotation and dilation transforms; the remainder give similar results) in Figures 3(a), 3(b), 4(a), and 4(b). For each figure, the tangent correlation is computed using *only* the tangent for the transform of interest (i.e. the rotation tangent when rotating the images, and the dilation tangent when dilating the images). The ρ_{cor} results are computed without



(a) Rotation by α radians.

(b) Dilation by a factor of $\alpha - 1$.

Fig. 5 KLT correlations ρ_{klt} between template (digit '6') and transformed digit images.

smoothing, and the $\rho_{\tan}$ results are computed using smoothing of the template and matched image. Although omitted from these figures in order to avoid excessive clutter, results were also computed for ρ_{cor} with smoothing of template and image, which was found to have significantly better transform invariance than ρ_{cor} without smoothing, but remained significantly inferior to $\rho_{\tan}$. In all of these figures, it is clear that $\rho_{\tan}$ provides much slower decay of correlation as the transform parameter is varied away from 0 (i.e. the identity transform), but at the expense of increased correlation against non-matching templates.

Since the proposed method is closely related to that of Uenohara and Kanade,[6] the behaviour of ρ_{klt} under rotation and scaling is of interest, and is displayed in Figures 5(a) and 5(b) respectively. For each transform, the template representation was constructed from a set of 91 transformed templates, the corresponding 91 transform parameters being selected by equally dividing up the displayed parameter range. The cumulative eigenvalue threshold ($\mu = 0.85$) from[6] was used, giving a representation on 10 and 7 eigenvectors for the rotation and dilation transforms respectively. As one would expect, the behaviour of ρ_{klt} is quite different to that of $\rho_{\tan}$, with a roughly constant, but lower correlation across the range of transform parameters, and without the clear peak, at $\alpha = 0$, with decaying correlation to either side.

These results indicate that the subspace methods do provide improved invariance to non-linear transforms, but it is not clear what their relative advantages are in actual template matching applications, or, in particular, what the associated trade-offs are when comparing increased correlation with a matching image region with increased correlations with non-matching regions. In order to make such an evaluation, we compare ROC curves for the different methods used to detect specific digits within a large collage image (see Figure 6) composed of 3000 digit images. These experiments provide a convenient means for evaluating the increased true match/false match trade-offs of each method, but are *not* intended to imply that the simple approach used here represents a practical

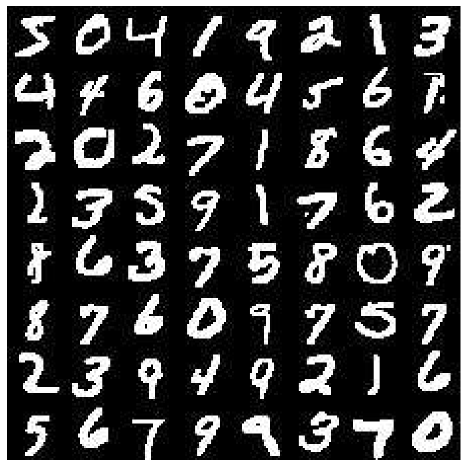
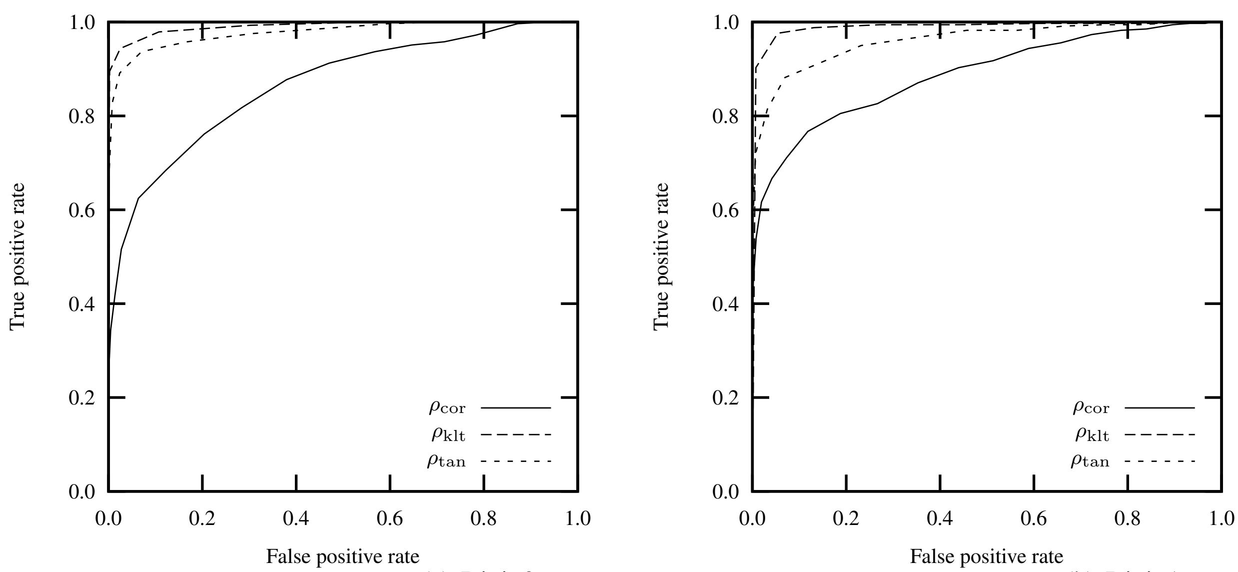
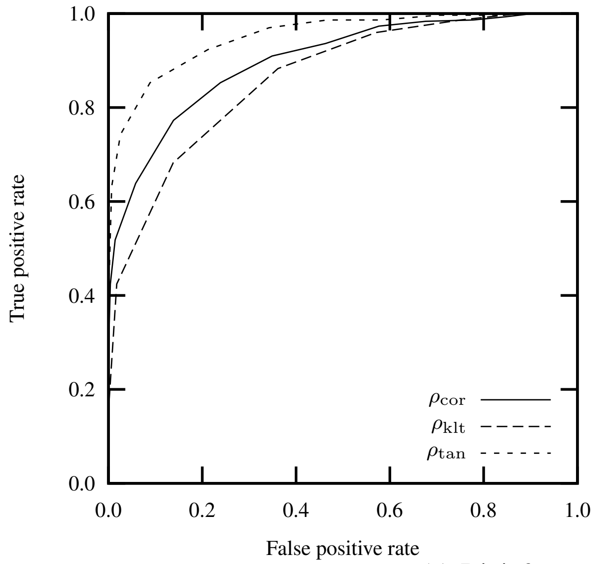


Fig. 6 A small section of the detection test image.

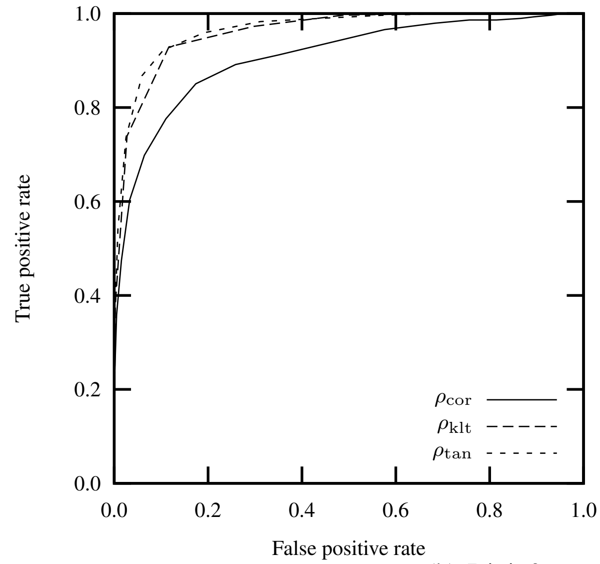


(a) Digit 0. (b) Digit 1.

Fig. 7 Comparison of ROC curves for detection of digits 0 and 1.

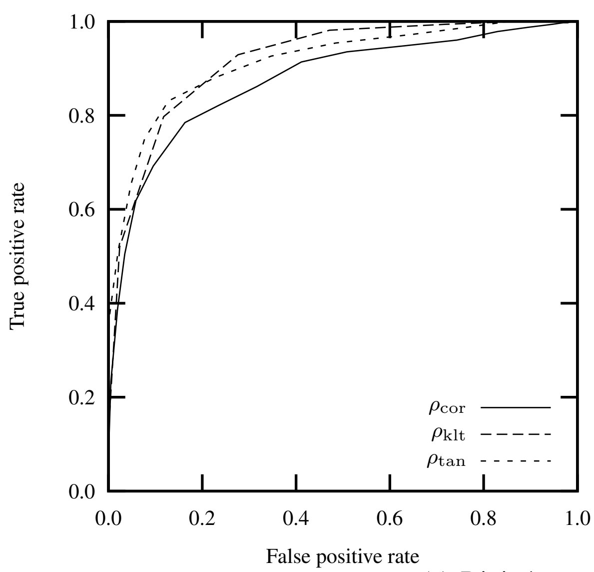


(a) Digit 2.

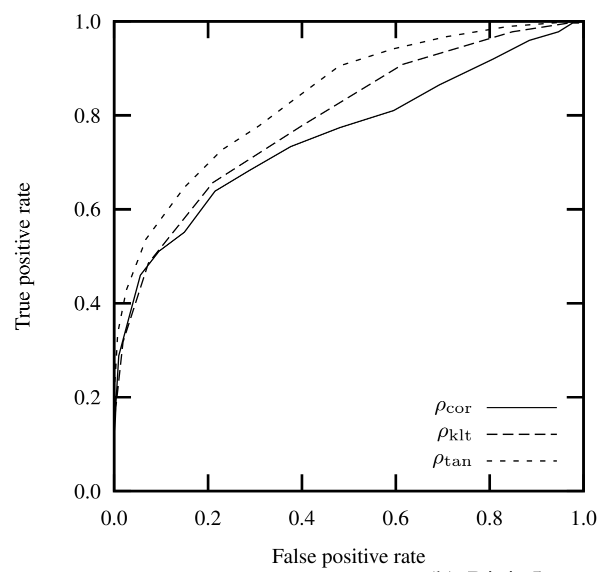


(b) Digit 3.

Fig. 8 Comparison of ROC curves for detection of digits 2 and 3.

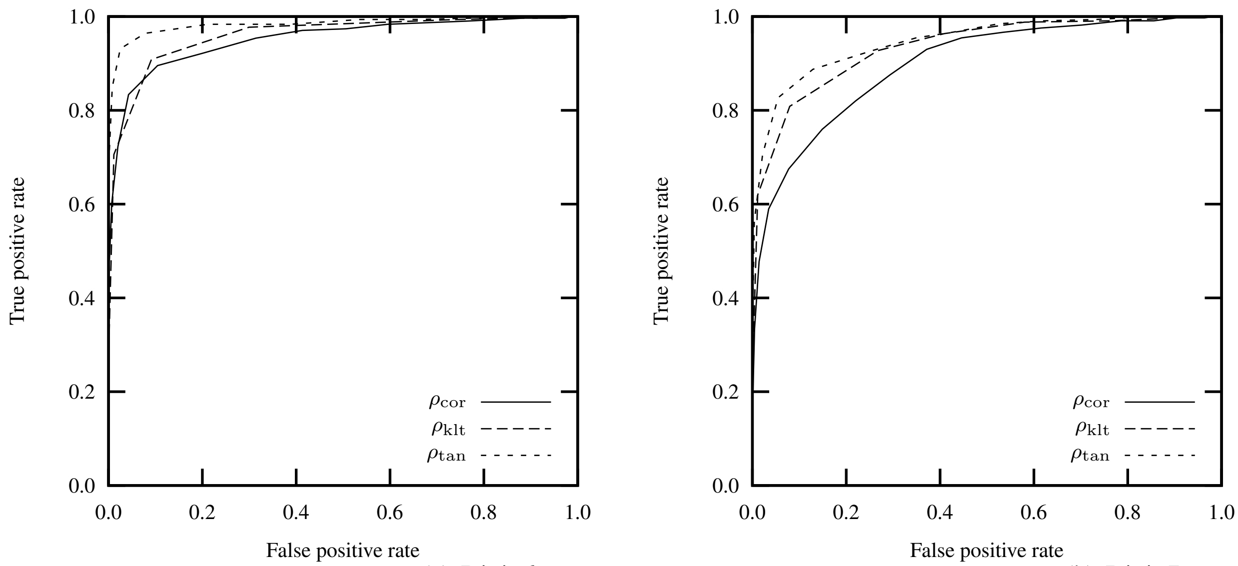


(a) Digit 4.



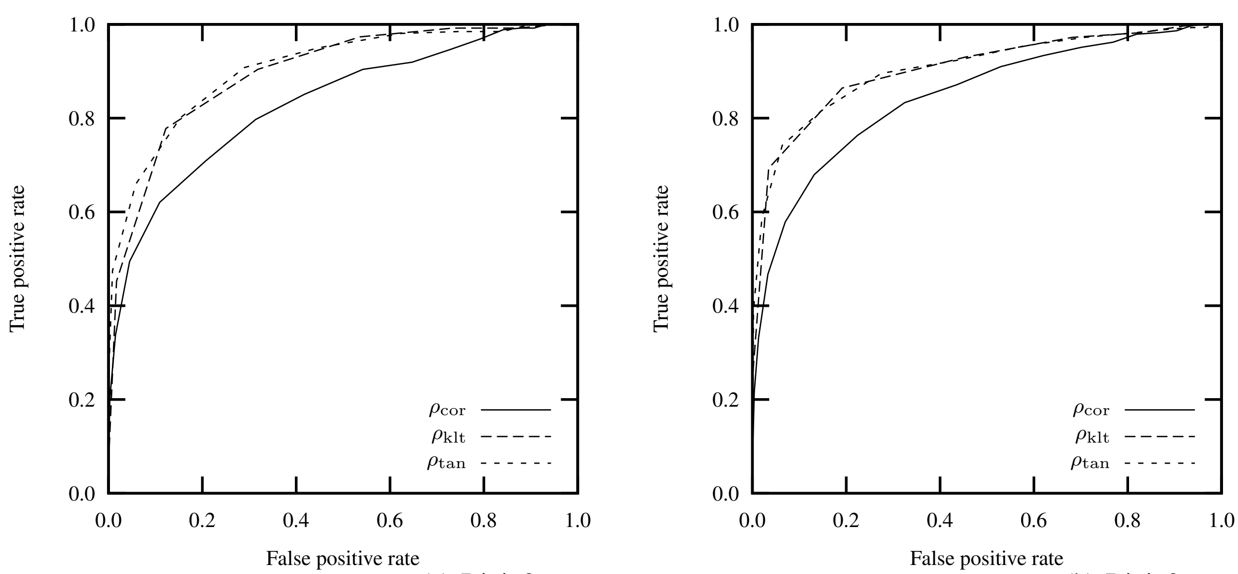
(b) Digit 5.

Fig. 9 Comparison of ROC curves for detection of digits 4 and 5.



(a) Digit 6.　(b) Digit 7.

Fig. 10 Comparison of ROC curves for detection of digits 6 and 7.



(a) Digit 8.　(b) Digit 9.

Fig. 11 Comparison of ROC curves for detection of digits 8 and 9.

digit detection algorithm. A digit is considered to have been detected in a given position in the collage if the correlation surface exceeds the detection threshold anywhere within that position in the collage. ROC curves (see Figures 7 to 11) are computed for the smoothed standard (i.e. ρ_{cor} with smoothing of template and image), smoothed KLT (i.e. ρ_{klt} with smoothing of template and image) and smoothed tangent correlations (i.e. ρ_{tan} with smoothing of template and image) by varying the range of the detection threshold.

Digit	ρ_{cor} (NS)	ρ_{cor} (S)	ρ_{klt} (NS)	ρ_{klt} (S)	ρ_{tan} (NS)	ρ_{tan} (S)
0	0.79	0.87	0.96	0.99	0.92	0.98
1	0.82	0.89	0.99	0.99	0.92	0.96
2	0.82	0.90	0.82	0.86	0.90	0.95
3	0.90	0.91	0.95	0.96	0.93	0.97
4	0.79	0.87	0.89	0.92	0.85	0.91
5	0.75	0.76	0.79	0.79	0.81	0.84
6	0.94	0.95	0.97	0.96	0.94	0.98
7	0.88	0.90	0.92	0.93	0.92	0.95
8	0.82	0.83	0.84	0.90	0.85	0.91
9	0.84	0.85	0.92	0.90	0.87	0.91
Mean	0.83	0.87	0.91	0.92	0.89	0.94

Tab. 1 Area under ROC curves for different correlation measures. Results for smoothed and unsmoothed templates and images are denoted by (S) and (NS) respectively.

The proposed correlation ρ_{tan} is superior to ρ_{cor} for all digits, by a significant margin in most cases. Where the advantage is small, we suspect it is related to the very large variations for these digits observed in the test data, which can obviously not be accounted for by the small transforms represented by the tangent correlation (significantly improved performance would probably be obtained by using multiple representative templates per digit). The correlation ρ_{klt} is superior to ρ_{tan} for digits “0” and “1”, approximately equivalent for digit “9”, and inferior for the remainder, in some cases quite significantly so, and is also inferior to ρ_{cor} for digits “2” and “8”.

In addition to the ROC curves, an overall numerical performance evaluation is provided by the areas under the ROC curves in Table 1. Smoothing improves all of the methods, including ρ_{klt}, which, with smoothing, is close in performance to ρ_{tan} with smoothing. For these results, the dimensionality of the KLT subspaces varied (over the set of 10 template digits) between 24 and 28 for the unsmoothed approach, and between 11 and 13 for the smoothed results. Reducing the subspace dimensionality to 6 (the same as for the proposed method) resulted in a large performance degradation, the areas under the ROC curves for ρ_{klt}, being 0.88 and 0.82 without and with smoothing respectively.

5 Conclusions

We have introduced a new technique for correlation based template matching with invariance to any chosen "small" transformations, with a computational cost which is a constant multiple of standard correlation based template matching. The advantages of this technique are shown empirically for a large test set of images of handwritten digits. This approach is closely related to the subspace matching approach of Uenohara and Kanade,[6] but differs in a number of important respects:

- The most significant difference lies in the construction of the subspace representation. The proposed technique gives more direct control of the subspace dimensionality, and in general, requires a lower dimensional subspace, corresponding to lower computational cost. The proposed method is also conceptually (and computationally) simpler, requiring computation of a single tangent vector for each transform with respect to which invariance is desired, rather than repeated application of each transform to construct a large set of examples.

- The ρ_{tan} correlation allows both positive and negative correlations, a distinction which is often important.

- The proposed method is more appropriate when invariance only to "small" transforms is explicitly required, giving a slow correlation decay as the transform parameter is varied from zero. The ρ_{tan} correlation includes a mechanism for preferring the primary template in the subspace, avoiding spurious matches with image blocks resembling differences between transforms of the primary template.

The proposed method has better performance than the KLT method, as measured by ROC curves in a detection application. When combining smoothing with the KLT method, the performance gap is smaller, but the proposed method has more convenient computation of the subspace, and lower computational requirements due to lower-dimensionality subspace.

While it is quite possible that the KLT method will provide superior performance in some applications, we argue that there are theoretical reasons for expecting superior performance (in the sense of increased correlation with transformed true matches while controlling the corresponding increase in false matches) when only attempting to use a subspace method to account for "small" non-linear transforms: as the transform parameters are allowed to become larger, the linear approximation, and therefore the subspace representation, becomes less accurate, and therefore less discriminating, with an increasing proportion of the subspace populated by vectors which are *not* transforms of the primary template, and therefore represent potential false matches with clutter in the images being matched.

A Product of Tangent Correlations ρ_0 and ρ_1

A relationship between the standard correlation measure ρ_{cor} and the tangent-based correlations ρ_0 and ρ_1 is easily derived:

$$\begin{aligned}
\rho_0(\mathbf{u},\mathbf{v})\rho_1(\mathbf{u},\mathbf{v}) &= \operatorname{sgn}(\langle \mathbf{v},\tilde{\mathbf{u}}\rangle)\frac{\|P\mathbf{v}\|}{\|R\mathbf{v}\|}\frac{|\langle \mathbf{v},\tilde{\mathbf{u}}\rangle|}{\|P\mathbf{v}\|} \\
&= \operatorname{sgn}(\langle \mathbf{v},\tilde{\mathbf{u}}\rangle)\frac{|\langle \mathbf{v},\tilde{\mathbf{u}}\rangle|}{\|R\mathbf{v}\|} \\
&= \frac{\langle \mathbf{v},\tilde{\mathbf{u}}\rangle}{\|R\mathbf{v}\|} = \frac{\langle \mathbf{v}, R\mathbf{u}/\|R\mathbf{u}\|\rangle}{\|R\mathbf{v}\|} \\
&= \frac{\langle \mathbf{v}, R\mathbf{u}\rangle}{\|R\mathbf{u}\|\|R\mathbf{v}\|} \\
&= \frac{\langle R\mathbf{v}, R\mathbf{u}\rangle}{\|R\mathbf{u}\|\|R\mathbf{v}\|} \quad (\text{since } \langle R\mathbf{v},\mathbf{1}\rangle = 0) \\
&= \rho_{\text{cor}}(\mathbf{u},\mathbf{v})
\end{aligned}$$

B Tangent Construction

In this appendix we provide a brief outline of the approach to computing the tangents for the dilation, rotation, parallel hyperbolic, and diagonal hyperbolic transforms. Greater detail, as well as the derivation of the tangent for the thickening transform, which is too lengthy for inclusion here, may be found in.[10]

Since we have only a sampled representation for our images, we compute the necessary derivatives by convolving our sampled representation with a Gaussian smoothing kernel, and making use of the result

$$f(x,y) = g(x,y)\star h(x,y) \quad \Rightarrow \quad \frac{\partial f}{\partial x} = \frac{\partial g}{\partial x}\star h \ \text{ and } \ \frac{\partial f}{\partial y} = \frac{\partial g}{\partial y}\star h,$$

allowing the transfer of the derivative to the smoothing kernel g.

B.1 *The Smoothing Kernel*

We use the smoothing kernel

$$g(x,y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right),$$

with derivatives

$$\begin{aligned}\frac{\partial g}{\partial x} &= -\frac{x}{\sigma^2} g(x,y) \\ \frac{\partial g}{\partial y} &= -\frac{y}{\sigma^2} g(x,y).\end{aligned}$$

B.2 *Dilation*

Define the operation of dilation by a factor of $1/(\alpha+1)$ as $T_\alpha f(x,y) = f((\alpha+1)x, (\alpha+1)y)$. We therefore have, after defining $u(x,\alpha) = (\alpha+1)x$ and $v(y,\alpha) = (\alpha+1)y$

$$\begin{aligned}\frac{d}{d\alpha} T_\alpha f(x,y) &= \frac{du}{d\alpha}\frac{\partial f}{\partial u} + \frac{dv}{d\alpha}\frac{\partial f}{\partial v} \\ &= x\frac{\partial f}{\partial u} + y\frac{\partial f}{\partial v} \\ &= x\left(\frac{\partial g}{\partial u} \star h\right) + y\left(\frac{\partial g}{\partial v} \star h\right),\end{aligned}$$

so that

$$\left.\frac{d}{d\alpha} T_\alpha f(x,y)\right|_{\alpha=0} = x\left(\frac{\partial g}{\partial x} \star h\right) + y\left(\frac{\partial g}{\partial y} \star h\right).$$

B.3 *Rotation*

Define the operation of anti-clockwise rotation by an angle of α radians as $T_\alpha f(x,y) = f(x\cos(\alpha) - y\sin(\alpha), x\sin(\alpha) + y\cos(\alpha))$. We therefore have, after defining $u(x,y,\alpha) = x\cos(\alpha) - y\sin(\alpha)$ and $v(x,y,\alpha) = x\sin(\alpha) + y\cos(\alpha)$

$$\begin{aligned}\frac{d}{d\alpha} T_\alpha f(x,y) &= \frac{du}{d\alpha}\frac{\partial f}{\partial u} + \frac{dv}{d\alpha}\frac{\partial f}{\partial v} \\ &= \left[-x\sin(\alpha) - y\cos(\alpha)\right]\frac{\partial f}{\partial u} + \left[x\cos(\alpha) - y\sin(\alpha)\right]\frac{\partial f}{\partial v} \\ &= -v\frac{\partial f}{\partial u} + u\frac{\partial f}{\partial v} \\ &= -v\left(\frac{\partial g}{\partial u} \star h\right) + u\left(\frac{\partial g}{\partial v} \star h\right),\end{aligned}$$

so that

$$\left.\frac{d}{d\alpha} T_\alpha f(x,y)\right|_{\alpha=0} = -y\left(\frac{\partial g}{\partial x} \star h\right) + x\left(\frac{\partial g}{\partial y} \star h\right).$$

B.4 *Parallel Hyperbolic Transform*

The parallel hyperbolic transform (see [10, Section IV.C]) with parameter α is $T_\alpha f(x,y) = f((1+\alpha)x, (1-\alpha)y)$. We therefore have, after defining $u(x,\alpha) = (1+\alpha)x$ and $v(y,\alpha) = (1-\alpha)y$

$$\begin{aligned}\frac{d}{d\alpha}T_\alpha f(x,y) &= \frac{du}{d\alpha}\frac{\partial f}{\partial u} + \frac{dv}{d\alpha}\frac{\partial f}{\partial v}\\ &= x\frac{\partial f}{\partial u} - y\frac{\partial f}{\partial v}\\ &= x\left(\frac{\partial g}{\partial u}\star h\right) - y\left(\frac{\partial g}{\partial v}\star h\right),\end{aligned}$$

so that

$$\left.\frac{d}{d\alpha}T_\alpha f(x,y)\right|_{\alpha=0} = x\left(\frac{\partial g}{\partial x}\star h\right) - y\left(\frac{\partial g}{\partial y}\star h\right).$$

B.5 *Diagonal Hyperbolic Transform*

The diagonal hyperbolic transform (see [10, Section IV.C]) with parameter α is $T_\alpha f(x,y) = f(x+\alpha y, \alpha x + y)$. We therefore have, after defining $u(x,y,\alpha) = x + \alpha y$ and $v(x,y,\alpha) = \alpha x + y$

$$\begin{aligned}\frac{d}{d\alpha}T_\alpha f(x,y) &= \frac{du}{d\alpha}\frac{\partial f}{\partial u} + \frac{dv}{d\alpha}\frac{\partial f}{\partial v}\\ &= y\frac{\partial f}{\partial u} + x\frac{\partial f}{\partial v}\\ &= y\left(\frac{\partial g}{\partial u}\star h\right) + x\left(\frac{\partial g}{\partial v}\star h\right),\end{aligned}$$

so that

$$\left.\frac{d}{d\alpha}T_\alpha f(x,y)\right|_{\alpha=0} = y\left(\frac{\partial g}{\partial x}\star h\right) + x\left(\frac{\partial g}{\partial y}\star h\right).$$

Acknowledgment

This work was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06NA25396 and was fully supported by the NNSA's Laboratory Directed Research and Development Program.

References

1. A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Upper Saddle River, NJ, USA, 1989.
2. W. K. Pratt, *Digital Image Processing*, John Wiley & Sons, New York, NY, USA, 2nd ed., 1991.
3. P. E. Anuta, "Spatial registration of multispectral and multitemporal digital imagery using Fast Fourier Transform techniques," *IEEE Transactions on Geoscience Electronics* **GE-8**, pp. 353–368, Oct. 1970.

4. J. Altmann and H. J. P. Reitböck, “A fast correlation method for scale- and translation-invariant pattern recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**, pp. 46–57, Jan. 1984.
5. A. Goshtasby, “Template matching in rotated images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-7**, pp. 338–344, May 1985.
6. M. Uenohara and T. Kanade, “Use of Fourier and Karhunen-Loeve decomposition for fast pattern matching with a large set of templates,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **9**(8), pp. 891– 898, 1997.
7. A. Pentland, B. Moghaddam, and T. Starner, “View-based and modular eigenspaces for face recognition,” pp. 84–91, June 1994.
8. B. V. K. V. Kumar, D. Casasent, and H. Murakami, “Principal-component imagery for statistical pattern recognition correlators,” *Optical Engineering* **21**(1), pp. 43–47, 1982.
9. R. Brunelli and T. Poggio, “Template matching: Matched spatial filters and beyond,” *Pattern Recognition* **30**, pp. 751–768, May 1997.
10. P. Y. Simard, Y. A. LeCun, J. S. Denker, and B. Victorri, “Transformation invariance in pattern recognition: Tangent distance and propagation,” *International Journal of Imaging Systems and Technology* **11**(3), pp. 181–197, 2000.
11. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE* **86**, pp. 2278–2324, November 1998.
12. Y. LeCun, “The MNIST database.” http://yann.lecun.com/exdb/mnist/.

Brendt Wohlberg received the BSc(Hons) degree in applied mathematics, and the MSc(Applied Science) and PhD degrees in electrical engineering from the University of Cape Town, South Africa, in 1990, 1993 and 1996 respectively. He is currently a technical staff member in the Mathematical Modeling and Analysis Group (T-7) at Los Alamos National Laboratory, Los Alamos, NM. His research interests include image coding, pattern recognition, wavelets and adaptive signal decompositions, and inverse problems in signal and image processing.

Kevin R. Vixie obtained a Ph.D. in 2001, under Andrew M. Fraser, while developing a research program at Los Alamos National Laboratory. He is currently a LANL technical staff member of the Mathematical modeling and Analysis group (T-7) in the Theoretical Division. His research interests range from image and signal analysis to variational methods and geometric measure theory. He is currently focusing on geometric analysis and development of physics informed regularizations and defensible metrics enabling principled comparisons of experimental data and simulations of these experiments.