

# Shaping Receptive fields for Affine Invariance\*

S. Ravela

Earth, Atmospheric and Planetary Sciences  
Massachusetts Institute of Technology  
Cambridge, MA, 02139  
ravela@mit.edu

## Abstract

*The Gaussian kernel has played a central role in multi-scale methods for feature extraction and matching. In this paper, a method for shaping the filter using the local image structure is presented. We propose an optimization formulation that densely estimates the filter’s affine parameters by minimizing an objective constructed from differential feature responses and seeks iterative, approximate solutions. A consequence of shaping the filters is affine invariance of the differential feature vector and it is shown that the shaped responses improve recognition performance.*

## 1 Introduction

Local continuum feature representations for recognition and retrieval are becoming increasingly common. The methodology employed in such techniques compares representations of images generated by using local features of the image brightness surface. Features obtained by applying such operators can equivalently be thought of as tunable spatial-frequency filters, statistical descriptors of the brightness surface, or approximations of its local shape.

We are specifically concerned with multi-scale differential features [2, 4, 5, 6, 12, 17, 19, 15, 20, 21]; a choice motivated by arguments [2, 5] that the local structure of an image can be represented in a stable and robust manner by the outputs of a set of multi-scale Gaussian derivative filters applied to an image.

These techniques commonly use the isotropic Gaussian scale-space, which precludes construction of affine invariant features. Whilst arguments have been made that isotropic scale-space features tend to behave well under deformations, it is difficult to ascertain just how well. Consider for example a pair of affine-deformed images with a single global affine transform, with no translation. That is,  $I_0(\underline{p}) = I_1(A\underline{p})$ . It is well known that the covariance congruence  $\Sigma_0 \Leftrightarrow A\Sigma_1A^T$  of the Gaussian filters at corresponding points produces equal

responses. Now, if isotropic filters are used to extract features, matching relies on “luck” rather than an appropriate adjustment of the norm. Such problems exist for any “unnormalized” filter function, such as Gabor filters used commonly in the literature [24, 10]. To be sure, techniques have been developed [9] that can match images under an affine distortions by deforming the filter to minimize the error energy between an image pair, but the construction of dense affine invariant continuum features from a single image has been elusive.

This paper presents a novel approach to adapt the shape of the filter using the local image structure, such that affine “invariance” is a natural consequence. Akin to producing rotational invariance by estimating local orientation in the image, or scale “invariance” by estimating scale from local image-structure, the proposed methodology operates in the affine scale-space and estimates the covariance (three parameters) to shape the filter. Shaping is local, and therefore this technique not only extends to images related by a global affine deformation, but ones that can be modeled by local affine variations as well. Although beyond the scope of this paper, the proposed method can also be used to shape other filter families. The shaping is also dense and shape parameters, if necessary, can be computed at every pixel location.

The proposed technique is an optimization approach. The objective takes the form

$$J[k] = - \sum_i [\mathcal{DI}(k - i, \Sigma_k)]^2 G(i, s\Sigma_k) \quad (1)$$

Here,  $J[k]$  is the objective at a pixel  $k$ ,  $\mathcal{DI}$  is a differential feature vector computed using a Gaussian at inner-scale  $\Sigma_k$  and the filter responses in a neighborhood are aggregated using a Gaussian  $G$  with scale  $s\Sigma_k$ , for a fixed  $s$ . That is, the outer scale is coupled to the inner scale. This formulation of the objective has its roots in the windowed second moment matrix [6] and related approaches to scale selection, where responses are aggregated in an outer-scale neighborhood. Here, instead

---

\*Supported in part by NSF grants 0100851 and ITR-0121182.

of relying on a single differential feature tuned to a fixed type of image structure, several differential features are used.

The initial guess for  $\Sigma_k$  is obtained using scale-selection with an isotropic scale-space (which in of itself provides invariance to a significant range of isotropic scale changes). Affine parameters are estimated by considering the partials of  $J$  with respect to the three parameters in  $\Sigma$ . This equation is solved by approximating the Jacobian and iterating to a solution using gradient descent. Upon convergence the features in  $\mathcal{DI}$  are affine invariant and, in general, the final filter parameters can be used to produce an affine invariant N-jet. The contribution of this paper is a method to use multiple differential features to synthesize shaped filters, densely if necessary.

These features are then applied for recognition. The plan for recognition is to compare estimated distributions of local features computed at multiple scales (multiples of the estimated filter-shape) Such an approach using differential features has been formulated earlier for unadapted filters [17, 20, 12, 1] and results suggest that the shape adapted representation outperforms one that doesn't.

## 2 Related Work

We are interested in factorizing “shape”, “content” and “noise” from images. Toward this end, differential features are used to represent content, scale-space is used to factor out “noise” and provide a framework to estimate “shape” with which differential features can be transformed to produce robust (tolerant to brightness and coordinate deformations) content features on a pixel by pixel basis. Feature statistics in windowed regions of the image are used as the image representation in recognition applications. With regard to this paper, there are two threads of related work. One that specifically pertains to shape adaptation and second, to the general framework. The second aspect is considered first, but please note that it is beyond the scope of this paper to review a large body of literature using this methodology. Instead, we focus on the community around differential features.

Gaussian multi-scale differential features and their representations have been studied in the literature. Methods have been developed using N-jets, steerability [3], “scale-shifts,” rotational invariants, and gray-scale normalizations to extract features, with tabulara approaches [15], angle constraints at interest points [21], distance constraints at sampled points [17], multidimensional histograms [20], shape index histograms [12, 1], concatenated multi-scale histograms of several differential features [17], and learned feature relationships [14] to represent images. The multi-scale

differential features used here can be related to commonly used texture features. In the context of image retrieval Ma et. al. [10] use Gabor filters to retrieve images with similar texture. Gabor jets [24] have also been used for face recognition. A comparison between Gaussian and Gabor filters, while instructive, is beyond the scope of this paper and a good review can be found in [17].

None of the above mentioned techniques deal with filters shaped with image structure. In particular, none of the techniques that rely on dense feature statistics to represent images produce affine invariant features. However, with regard to shaping, there is a significant body of emerging literature. Almost all of these techniques examine affine invariance in the context of the second moment matrix, and a majority for features at interest points. This includes Lindeberg and Garding [8] and Mikolajczyk and Schmid [11]. The proposed algorithm is different in the following ways. First, it must be viewed in the context of the use for dense representations where, by definition, shape estimates must be computed densely. Second, there is no reliance on the second moment matrix or, for that matter, any single feature. The motivation here is to use an ensemble of features, responding to different structures in the image *simultaneously* to estimate the filter parameters. Third, the optimization framework proposed here estimates the filter parameters by iteratively deforming the filter. Finally, the ensemble of features used for estimating filter-shape can also be used as image features and therefore “invariant” features are produced at a point along with the final filter shape.

## 3 Dense Affine Invariant Features

The first step in constructing a differential feature representation is computing a vector of spatial image derivatives. For example, given an image  $I$ , the first two orders of spatial derivatives can be used as a feature (vector). This vector approximates the shape of the local intensity surface in the sense of a second order Taylor approximation. Including higher orders produces a more precise approximation. Derivatives capture useful statistical information about the image. The first derivatives represent the gradient or “edgeness” of the intensity and the second derivatives can be used to represent curvatures (e.g. bars, blobs).

### 3.1 Affine Gaussian scale-space:

However, it is important that derivatives be computed in a stable manner. Derivatives will be stable if, instead of using just finite differences, they are computed by filtering an image with normalized Gaussian derivative filters (actually any  $C^\infty$  function will do [2]). In two dimensions, a Gaussian derivative is the deriva-

tive of the function

$$G(\cdot, \Sigma) = \frac{1}{2\pi|\Sigma|} e^{-\frac{1}{2}\underline{p}^T \Sigma^{-1} \underline{p}}$$

In the frequency domain, a Gaussian derivative filter is a band-pass filter. Computing derivatives by filtering with a Gaussian derivative at a certain scale, therefore, implies that only a limited band of frequencies are being observed. Thus, in order to describe the original image more completely, a multi-scale representation is necessary. Sampling the scale-space of the image becomes essential.

It has been shown by several authors [2, 4, 6, 19, 25] that under certain general constraints, the Gaussian filter forms a unique operator for representing an image across the space of scales. It is beyond the scope of this document to engage in a full discussion about the scale-space image representation and, instead, the reader is referred to the following papers [2, 6, 19, 25].

### 3.2 Filter Shaping

The central question is what scales to select and how. Here, scale refers to the three parameters of the covariance matrix  $\Sigma$  in the two dimensional case. To answer this question we start with a one dimensional example, where there is only one parameter to estimate, and then extend the results (trivially) to two dimensions. Let  $\delta^n I$  be a differential version of the one dimensional signal  $I$ . We seek to optimize the objective, at a given point  $k$ :

$$\begin{aligned} J[k] &= - \sum_i [\sigma_k^n \delta^n I \otimes g(\cdot, \sigma_k)]_{k-i}^2 g(i, s\sigma_k) \quad (2) \\ &\doteq - \sum_i L_{k-i} g(i, s\sigma_k) \quad (3) \end{aligned}$$

Here  $\sigma_k$  is the inner scale associated with pixel  $k$ , and  $s\sigma_k$  is the outer scale (see [6]), and  $g$  is the normalized one dimensional Gaussian. The symbol  $\otimes$  refers to convolution and  $s$  is a fixed constant. The gradient of the objective can be written

$$\frac{dJ[k]}{d\sigma_k} = - \sum_i \left[ \frac{dL_{k-i}}{d\sigma_k} g(i, s\sigma_k) + L_{k-i} \frac{dg(i, s\sigma_k)}{d\sigma_k} \right] \quad (4)$$

The solution is obtained using gradient descent and makes the following approximation. The first part of the Jacobian  $\frac{dL_{k-i}}{d\sigma_k} g(i, s\sigma_k)$ , is dropped. The rationale is that one of the two terms produced by differentiating the first part is similar to the second part in the Jacobian, and the second one is a higher order term involving  $\nabla^2 g(i, \sigma_k)$ . Since the outer scale and inner scale have been coupled and changes are propagated to  $L_{k-i}$  at every iteration of the numerical solution to Equation 4,

extra computations of terms whose contributions can be subsumed by the outer scale adjustment can be avoided.

Gradient descent is used to recover  $\hat{\sigma}_k$ , the final scale. That is,  $\sigma_k^{t+1} = \sigma_k^t - \alpha \frac{dJ^\#}{d\sigma_k}$ , where  $dJ^\#$  is the approximated Jacobian. During each iteration the Jacobian is held constant at the previous scale estimate, the scale is updated, and the Jacobian (including inner-scale filtering) is computed with the new estimate. This process is repeated till convergence (or an iteration limit).

In Figure 1, a discretized cosine wave and the estimated scale is shown. The sine wave is assumed to be the differential signal and the scale of the Gaussian is in good agreement with the scale of the local structure (the negative lobe).

This procedure is extended to two dimensions. Here, the estimated parameters are  $\Sigma^{-1}$ , say  $a_{11}$ ,  $a_{12}$ , and  $a_{22}$  and during each iteration of the descent, the component Jacobians are held fixed from the previous iteration.

There are however three considerations left to be addressed. First, the initial guess for the parameters is obtained using the scale-selection method outlined in [7]. Second, the growth and shrinkage of the parameters are bounded and we usually set  $\sigma_{max}$  to be a proportion of the image size, and  $\sigma_{min} = 0.6$ , determined from the eigen values of  $\Sigma$ .

The third issue pertains to the choice of differential features that make up  $L$ . Since  $L$  is not differentiated (but it is re-evaluated at every iteration), therefore there is considerable flexibility for designing the forcing terms. In general, we use the form  $L_j = (F_j / \|F\|)^2$  and consider a few choices for  $F_j$ .

One choice that has had some success is the single feature  $([I_{uu} + I_{vv}]_j)$ , ( $u, v$  are local directions along the major and minor axes of the evolving filter and the cartesian basis in the isotropic case). Such objectives however are tuned to specific features and we suggest the use of a mixture of differential features.

With regard to using multiple differential features, another choice is the norm of lower order feature vector,  $D_j = [\{\sqrt{I_u^2 + I_v^2}\}_j \{I_{uu} + I_{vv}\}_j]^T$ , computed using the evolving filter. This takes the form  $L_j = D_j^T W_D D_j$ , where weight matrix  $W_D$  is pre-specified, also see [7].

We have had considerable success using a third form of  $F$ , where  $F_j = \det(C_{DD}^{(j)})$ , the determinant of the covariance matrix of a feature vector comprising of rotational invariants (to order two), computed in a window of size proportional to the evolving filter parameters and centered at  $j$ .

In Figure 2, the results of shape adaptation are shown. The ellipses are drawn over selected points and are a multiple (half power) of the estimated parameters of the outer-scale filter. Differential features computed

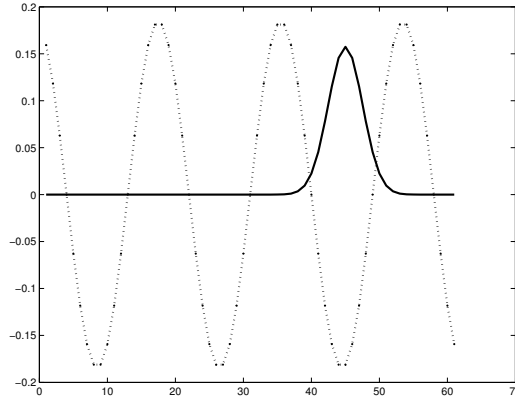


Figure 1: This figure demonstrates scale adaptation using the algorithm developed in this paper. The dotted line is the original signal and the solid curve is the shape of the final filter estimated at a point on the x-axis, directly under the maximum value of the filter. the initial filter was  $\sigma = 1$ , the final is  $\sigma = 2.5360$

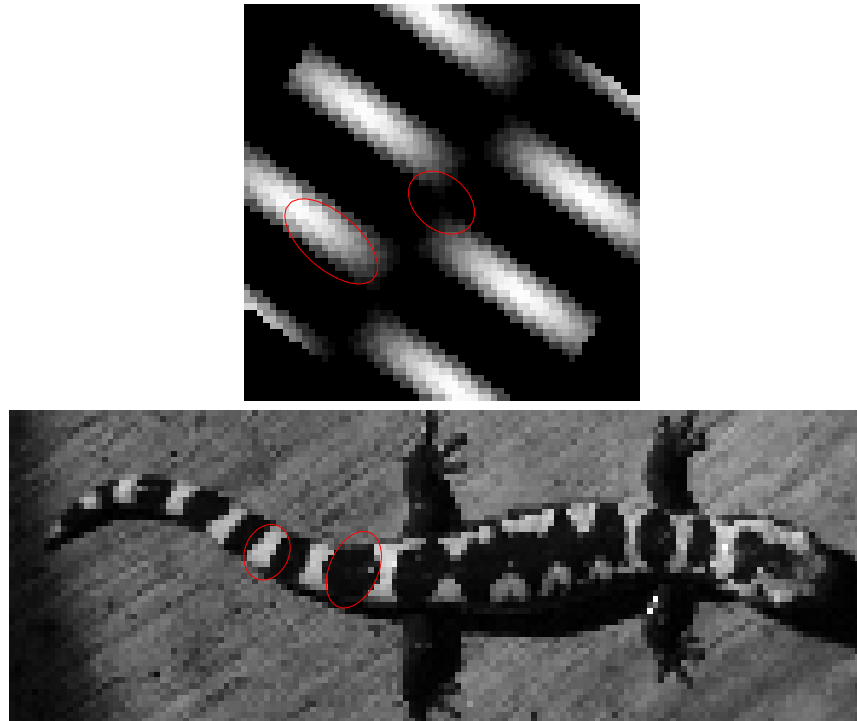


Figure 2: This figure demonstrates shape adaptation using the algorithm developed in this paper. The ellipses are half power contours.

using these filters at the indicated locations are invariant to affine changes.

#### 4 Application to Recognition

There are several applications of shape adapted differential feature representations. In particular, we are interested in their use in recognition/retrieval applications. The steps involved in deducing similarity between a “query” and database image are as follows: Database images are filtered *a priori*. The isotropic

scale is estimated at uniformly sampled locations in the image and used as a first guess to shape the filters. At scalar multiples  $[\frac{1}{2}, \frac{1}{\sqrt{2}} \dots 2]$  of estimated shape parameters, the first and second order shape adapted N-Jet are computed and normalized for brightness changes [15, 21, 17]. A query image is processed the same way. Histograms of the multi-scale differential features, one histogram per feature per scale, are concatenated into a vector and correlated to measure sim-

ilarity. The top  $N$  ranks are presented to the user with the objective of locating the query in the database. We have used this methodology before with unadapted differential features [17] and this algorithm adds the shape-adaptation component. We call this the SFH-1 (shaped feature histogram-1D) algorithm.

We demonstrate the use SFH-1 on two recognition tasks. The first is face recognition using the ORL set, and the second is salamander recognition. The performance of the unadapted version of the features with the same representation and similar algorithm, called CO1, is well documented on these tasks [18, 16] and the reader is referred to these papers to obtain details on the databases, protocols, and methods used.

The marbled salamander is an endangered species in Massachusetts and extensive studies are being conducted at this time to study their migratory patterns. Marbled salamanders have textured backs and, we have suggested that these can be used to recognize them, individually. Pictures of salamanders are obtained from the field and placed in a database. The database used for obtaining the salamander was collected over a four year period (99-02), from fourteen different sites (totaling approximately 3000). Multiple images of salamanders may be taken at a single time and other instances of the same individual may be obtained at different sites and different times. A total of 69 queries were used to test the algorithm on a small subset. The idea is to see if a newly imaged salamander is present in the database. From a vision point of view this is challenging because salamanders have flexible shapes because their morphology changes with time.

For the salamander recognition task a leave one out test was used to evaluate the performance of the algorithm. The objective specification is to provide at most 5 retrievals and seek 100% recall. It is beyond the scope of the paper to discuss the parameters used in CO1, but the interested reader can refer to [16]. For SFH-1, five scales around the estimated shape parameters were used. The shape parameters were estimated at every 10th pixel and interpolated for in-between pixels. The performance of the algorithm is depicted in Table 1, where we see that approximately 95% recognition rate is achieved. This is an improvement over the un-adapted version.

The ORL (Olivetti Research Lab) collection is a publicly available collection of 400 faces. This collection contains 40 individuals. The database contains small view, gesture, and intensity variation. The evaluation methodology follows the one described by Sim et. al. [22]. During each trial a database is randomly split into a training set and a test set. The configurations of training set per trial uses 5 exemplars per per-

Technique	ORL
CMU [22]	97%
CO1 [18]	95%
Eigen-face [23]	95%
SFH-1	98%

Table 2: The performance of SFH-1 method in relation to other techniques.

son. The remaining faces for the person become the test set. Each of these test set images becomes a query. A query is matched with all of the training set and the identity of the best matching training set image is ascribed to the query. Over a large (100) number of trials the proportion of correctly identified people is reported as the recognition rate. For example, in the ORL set a trial will consist of 200 training and test images each. Thus, over 100 trials 20,000 queries (test set) are matched with a random training/test pick at every trial. The parameters used for SFH-1 are the same as for the previous test, and the comparable CO1 parameters are shown in [18]. Results in Table 4 indicate a performance improvement over the unadapted version, Eigenfaces and Sim’s method [22].

## 5 Summary and Conclusions

The principal contribution of this paper is a new algorithm for shape adapted filtering, exploring the affine Gaussian scale space. These results extend the technique for recognition using differential features – in particular receptive field histogram framework. The proposed technique can be used to estimate filter-shape parameters densely. It does not rely on any single feature detector (or interest operator), but uses an ensemble to compute the best shape parameter. The exploration of the affine scale-space is posed as a minimization problem, and in particular the approximations presented in the paper can be used to effectively estimate parameters. The algorithm presented here is applied to recognition in the salamander and face recognition tasks and observe that there is some improvement over using “unadapted” filters. These results, however, are somewhat preliminary because the gains in recognition performance in these tasks are not very large. What is clear however is that this methodology can be used for matching in several applications to establish correspondence, across motion and stereo for example, or even feature correspondence based methods for recognition. This paper also indicates the next steps to be taken. In particular, we are interested in developing efficient interpolation schemes so that affine shape parameters need not be computed everywhere. We are also interested in developing filters that can adapt to nonlinear shape de-

Algorithm	Rate at Rank	1	2	3	4	5
COI		50/69	54/69	60/69	61/69	63/69
SFH-1		55/69	59/59	62/69	64/69	66/69

Table 1: The performance of SFH-1 method on Salamander Recognition

formations.

## References

- [1] Chitra Dorai and Anil Jain, "COSMOS - A representation scheme for free form surfaces", ICCV 95, pp. 1024-1029, 1995.
- [2] L M J Florack, The Syntactic Structure of Scalar Images, PhD Dissertation, University of Utrecht, 1993
- [3] W. T. Freeman and E. H. Adelson, The design and use of steerable filters, IEEE Trans. Patt. Anal. and Mach. Intel., 13(9):891-906, 1991
- [4] J. J. Koenderink, The Structure of Images, Biological Cybernetics, 50:363-396, 1984.
- [5] J. J. Koenderink and A. J. van Doorn, Representation of Local Geometry in the Visual System, Biological Cybernetics, 55:367-375, 1987
- [6] T. Lindeberg, Scale-Space Theory in Computer Vision, Kluwer Academic Publishers, 1994
- [7] T. Lindeberg, Feature detection with automatic scale selection, International Journal of Computer Vision, 30(2):79-116, 1998
- [8] T. Lindeberg and J. Garding, Shape-adapted smoothing in estimation of 3D-shape cues from affine deformations of local 2D structure, Image and Vision Computing, 15(6):415-434, 1997
- [9] R. Manmatha, Matching Affine-Distorted Images, Ph.D. Dissertation, University of Massachusetts at Amherst, 1997
- [10] W. Y. Ma and B. S. Manjunath, Texture-Based Pattern Retrieval from Image Databases, Multimedia Tools and Applications, 2(1):35-51, Jan. 1996
- [11] K. Mikolajczyk and C. Schmid, An affine invariant interest point detector. European Conference on Computer Vision, vol. 1, 128-142, 2002
- [12] C. Nastar, The image shape spectrum for image retrieval, Technical Report 3206, INRIA, June 1997.
- [13] Olivetti Research Labs, Face Dataset, [http : //www.cam - orl.co.uk/facedatabase.html](http://www.cam-orl.co.uk/facedatabase.html)
- [14] J. Piater, Visual Feature Learning, PhD Dissertation, University of Massachusetts at Amherst, 2001.
- [15] Rajesh Rao and Dana Ballard, Object Indexing Using an Iconic Sparse Distributed Memory, Proc. International Conference on Computer Vision, pp. 24-31, 1995.
- [16] S. Ravela and R. Gamble, On Recognizing Individual Salamanders, Proc. Asian Conference on Computer Vision, Jeju Island, Korea, 2004.
- [17] S. Ravela, On multi-scale differential features and their representations for recognition and retrieval, PhD Dissertation, University of Massachusetts at Amherst, 2003
- [18] S. Ravela and A. Hanson, On multi-scale differential features for face recognition, In Proc. Vision Interface 01, Ottawa, June 2001.
- [19] B. M. ter Har Romeny, Geometry Driven Diffusion in Computer Vision, Kluwer Academic Publishers, 1994
- [20] Bernt Schiele and James L. Crowley, Object Recognition Using Multidimensional Receptive Field Histograms, Proc. 4th European Conf. Computer Vision, Cambridge, U.K., April 96.
- [21] Schmid, R. Mohr, Local Grayvalue Invariants for Image Retrieval, PAMI (19), No. 5, pp. 530-535, May 1997.
- [22] T. Sim, R. Sukthankar, M. Mullin, and S. Baluja, High-Performance Memory-based Face Recognition for Visitor Identification, 1999 (see [http : //www.ri.cmu.edu/pubs/pubs2772.html](http://www.ri.cmu.edu/pubs/pubs2772.html) )
- [23] M. Turk and A. Pentland, Eigen Faces for Recognition, Jnl. Cognitive Neuroscience, 3:71-86, 1991.
- [24] L. Wiskott, J.-M. Fellous, N. Kruger and C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans Patt. Anal. and Mach. Intell. 17(7):775-779, 1997.
- [25] A. P. Witkin, Scale-Space Filtering, Proc. Intl. Joint Conf. Art. Intell., pp. 1019-1023, 1983