Appearance based recognition using spatial and discriminant influence

Qi Li
Department of Mathematics & Computer Science
Western Kentucky University
Bowling Green, USA
qi.li@wku.edu

Chang-Tien Lu
Department of Computer Science
Virginia Tech
Falls Church, USA
ctlu@vt.edu

***Abstract*—Appearances of objects lie in high-dimensional spaces. For a given recognition task, feature selection aims to select most effective features in order to reduce the recognition cost and improve recognition accuracy. Feature selection can be achieved by a bottom-up scheme, e.g., using spatial information, or a top-down scheme, e.g., using class information. In this paper, we propose a model to integrate spatial and discriminant influence for appearance based recognition, where locality oriented Fisher score is introduced to estimate the discriminant influence. We use Lipschitz regularity to construct image representation. We present a case study of embryo stage recognition to test the performance of the proposed method. We also obtain new insights on the comparison between spatial and discriminant influence.**

***Keywords*-Appearance based recognition; feature selection; spatial influence; discriminant influence;**

I. Introduction

Appearances of objects lie in high-dimensional spaces. For a given recognition task, feature selection aims to select most effective features in order to reduce the recognition cost and improve recognition accuracy. Feature selection can be achieved by a bottom-up scheme, e.g., using spatial information [7], [10], [12], [18]. For example, Harris detector [7] uses gradient auto-correlation to define spatial influence. Bottom-up scheme aims to output features (interest points) repeatable across different image/illumination transformations, which helps construct robust and compact representation of image data. Bottom-up scheme has a wide range of applications, such as object recognition [8], image retrieval [13]. Bottom-up feature selection is an important step to build a generative model for object recognition [19], [3]. A generative model is basically a graph model with a relatively small number of features [8]. Generative models are strong in addressing "weak-alignment" recognition tasks where the shapes of different objects contain significant variations.

Feature selection can also be achieved by a top-down scheme, e.g., using class or context information [5]. Gao and Vasconcelos [5] argued that spatial information (such as edge, corners) may not always reveal good saliency of visual objects, and thus proposed a discriminant top-down selection method for visual recognition, where the discriminancy is computed via maximum marginal diversity [16].

Recently, integration of bottom-up and top-down feature selection received attention in the area of visual classification, including object detection [14] and object recognition [8]. (Object detection is a binary classification problem.) To speedup object detection, Navalpakkam and Itti [14] proposed a model to integrate bottom-up and top-down attention, where the top-down component uses accumulated statistical knowledge of the visual features of the desired search target and background clutter, to optimally tune the bottom-up maps such that target detection speed is maximized. In the context of object recognition, Holub and Perona [8] proposed a model to combine the generative model and Fisher kernels, which brings considerable improvement of the performance of generative models.

In this paper, we propose a model to integrate bottom-up and top-down feature selection for appearance-based recognition. A motivation of our study comes from the insight that the appearances of certain objects are weakly-textured, and a few of features may not be robust with respect to the imaging variations, such as illuminations or even image noise. Thus, we explore the opportunity of using relatively large numbers of features (in the magnitude of hundreds), compensated with an assumption that images requires alignment in our study. We introduce locality oriented Fisher scores to estimate the top-down influence, where locality is characterized by wavelets. For the robustness with respect to illumination, we use Lipschitz regularity based representation.

We present a case study to show the effectiveness of the integrated model. The case study is on the recognition of developmental stage of an embryo based on gene expression pattern images [9], [6], which is an important step towards gene expression analysis. The study convinces us the effectiveness of the integration of spatial (bottom-up) influence and discriminant (top-down) influence in selecting good features for appearance based recognition. The studies also bring us new insight on the comparison between bottom-up and top-down feature selection schemes.

The rest of the paper is organized as follows: Section II gives related work. Section III introduces locality oriented Fisher discriminant scores. Section IV proposes an integrated model, and Lipschitz regularity regularity based

representation. Section V presents three case studies. Finally, conclusion is given in Section VI.

II. Related work

Walther *et al.* [18] proposed a bottom-up model for selective attention, where bottom-up saliency map is contributed by the color feature maps, intensity feature maps, and orientation feature maps. They showed the proposed bottom-up visual attention can strongly improve learning and recognition performance in the presence of large amounts of clutter.

Vasconcelos [16] proposed a discriminant feature selection via maximization of marginal diversity (MMD); for multi-class problems, one-verse-all strategy is applied. N. Vasconcelos and M. Vasconcelos [17] proposed an information theoretic feature selection to achieve a good balance between maximizing the discriminant power of selected (local) features and minimizing their redundancy. The method is tested on image retrieval, where the comparison between two images is achieved by the comparison of Gaussian mixtures of the compact sets of discriminant (local) features detected from the images. Gao and Vasconcelos [5] presented a discriminant saliency method, based on MMD [16], to detect visual objects from cluttered backgrounds.

Navalpakkam and Itti [14] proposed a SNR based model to integrate bottom-up and top-down attention for optimizing detection speed, where SNR (signal-noise-ratio) characterizes the discriminant ratio of spatial influence of target objects over the spatial influence of distractors. Navalpakkam and Itti showed the model, with little computational cost in the form of multiplicative top-down gains on bottom-up saliency maps, predicts many reported bottom-up or top-down influence on human visual search behavior.

Holub and Perona [8] proposed a model to combine generative model and Fisher kernels for object recognition. The generative model used in [8] is a constellation model that aims to find optimal appearance and shape parameters $\{\theta_a, \theta_s\}$ during the mapping of interest points to model parts. A Fisher kernel is a gram matrix constructed by "Fisher score" feature that is the derivative of log likelihood of the parameters of a generative model. (Note that Fisher score used in [8] is different from Fisher criterion score used in this paper.)

Fisher criterion score [1] has been widely used for feature selection. In [4], Fisher criterion score is used to select most discriminant features of microarray expression data and achieved substantial improvement of recognition accuracy.

III. Locality oriented Fisher score

Fisher score was proposed to maximize the ratio of between-class variation over within-class variation. More specifically, given an attribute p, its Fisher criterion score is defined as follows:

$$\text{score}(p) = \frac{\sum_c |v_c(p) - v_t(p)|^2}{\sum_c \sum_{j \in c} |v_j(p) - v_c(p)|^2}, \tag{1}$$

where c is a class label, v_c is the mean of all training instances of attribute p in class c, v_t is the (total) mean of all instances of attribute p, and v_j is the instance of j-th training item. The most discriminant attribute is assigned by the highest Fisher score. Thus by sorting attributes according to their Fisher scores, a number of most discriminant attributes contribute a good feature vector for recognition, e.g., the use of nearest neighbor under Euclidean distance as a classifier. The number of most discriminant attributes is usually determined via cross-validation.

We introduce locality oriented Fisher scores to estimate discriminant influence where the locality is captured by wavelets. This score aims to maximize the ratio of local between-class variation and local within-class variation:

$$D(p) = \frac{\sum_c |g_c(p) - g_t(p)|^l}{\sum_c \sum_{j \in c} |g_j(p) - g_c(p)|^l}, \tag{2}$$

where l is a positive number, c is a class label, $g_j(p)$ is the wavelet coefficient of j-th training instance (i.e., the convolution of the neighborhood of p in j-th training instance with a wavelet filter), $g_c(p)$ is the mean of the wavelet coefficients of training instances in class c, and $g_t(p)$ is the overall mean of wavelet coefficients in p.

It is known that wavelets have several desirable properties: compact supports, symmetry, and/or high-vanishing moments, orthogonality, etc. Given a wavelet ψ (for simplicity, let us assume it is on $\mathbb{R}$), *compact support* indicates $\psi(x) \equiv 0$ out of some finite interval; *symmetry* indicates $\psi(x_0 - x) = \psi(x)$, for some $x_0 \in \mathbb{R}$; *vanishing moment* k indicates $\int x^l \psi(x) dx, l = 0, \ldots, k$; *orthognality* indicates $\int \psi(x)\psi(x - j)dx = 0, \forall j \in \mathbb{Z}$. Compact support is the key property for a wavelet technique to perform the local analysis. Vanishing moment is also a useful property for local analysis. Note that if a local region is smooth, it can be approximated by some low-order polynomials. Convolving with a wavelet of some-degree vanishing moment, its associated wavelet coefficients are small. Thus the magnitude of wavelet coefficients can characterize the smoothness of a local region. The work in signal or visual processing has found the importance of symmetry. Orthognality may be arguable depending on what space the data lies in. If the data is in L^2, it is desirable; Otherwise, it may be worthless.

We will use least asymmetric Daubechies wavelet to capture the locality in determining the Fisher criterion score. The least asymmetric Daubechies wavelet is constructed by constraining the phase of the so-called transfer function as close to linear as possible. (More details can be found in Chapter 8 of [2].)

It is worth noting that in our Fisher score formulation, we introduce the norm parameter l. In standard Fisher score, l

is always fixed as 2, i.e., Euclidean norm. It is known that in resisting outlier attributes, Euclidean norm may not perform best. In the later case study, we will observe the value of this generalization.

IV. Integrating spatial and discriminant influences

Denote p as an image point, J a set of training data, and i the index of a certain spatial filter such as Gradient auto-correlation [7], [15], Laplacian [12], and DoG [10]. Denote $\{S_j^i\}_{j\in J}$ as the spatial influence maps of all training images associated with a certain spatial filter. Denote T^0 as the unsupervised operator $T^0(\{S_j^i(p)\}_{j\in J}) = \frac{1}{|J|}\sum_{j\in J} S_j^i(p)$, which gives bottom-up feature selection. Denote $T^{i>0}$ as a supervised operator, such as locality oriented Fisher score with a certain norm l, which gives top-down feature selection.

Our model integrates a set of unsupervised and supervised operators that are applied to a set of spatial influence maps as follows:

$$\begin{aligned}\text{influence}(p) &= \sum_{k,i} \alpha_{k,i} T^k(\{S_j^i(p)\}_{j\in J}),\\ &\text{subject to } \sum_{k,i}\alpha_{k,i} = 1,\end{aligned}$$

where the weight parameters $\alpha_{k,i}$ reveal the prior of different bottom-up and top-down influences in a specific appearance based recognition task. The weight parameters can be learned by applying cross-validation to training data, i.e., optimal weights are decided by the recognition accuracy on validation data.

After each image point is assigned with a certain influence value, best features can be selected according to the order of their influence. Fig. 1 shows the influence maps of embryo images overlaid by fifty best feature points (i.e., pixels of strongest influence), illustrating the integration of two popular spatial influences— gradient auto-correlation and Laplacian—with discriminant influence, respectively. We can observe that feature points under gradient auto-correlation influence spread in the entire embryo plane with any specific concentration, and feature points under integrated influence have better concentration. (The higher recognition accuracy achieved by integrated features, shown in later experiments, explains the value of the concentration.)

With a set of feature points P, we can construct feature vectors (compact image representations) for appearance based recognition. A convenient and efficient way for constructing feature vectors is to use the intensities of those feature points, i.e., $\{I(p)\}_{p\in P}$. In the following, we have a comparison among the linear separability of these feature vectors where the feature points are selected via spatial, discriminant and integrated influence, respectively. (Note that linear separability is desirable to support efficient classifiers.) We use embryo images as examples, and apply Linear Discriminant Analysis (LDA) to visualize the feature vectors in 2-D plane. The dimension of embryo image is 320×128. Our data contains three classes (leading to two-dimensional LDA space). We will show a PCA+LDA representation as a comparison. Fig. 2 shows four different LDA representation. The first two classes of embryo data are shown for the clarity of comparison of the representation. The bold labels indicate the data items violating linear separability. From Fig. 2, we can see that the integrated influence contributes to feature vectors of best linear separability. This example gives us an insight of the effectiveness of integrating spatial and discriminant influences in improving the linear separability of the image representation.

We introduce Lipschitz regularity based feature vectors to improve the robustness with respect to illuminations. An image I is pointwise Lipschitz $\alpha \geq 0$ at p, if there exists $K > 0$ and a polynomial f_p of degree $m = \lfloor\alpha\rfloor$ such that $|I(q) - f_p(q)| \leq K\|q-p\|^\alpha, \forall q$. Furthermore, if I has a Lipschitz α regularity at P that is isolated and non-oscillating, it is uniformly Lipschitz α in the neighborhood of p. Mallat and Zhong [11] gave a method to estimate the Lipschitz regularity of an image point p by the decay of associated wavelet coefficients across scales. More specifically, they perform a regression on the following formula:

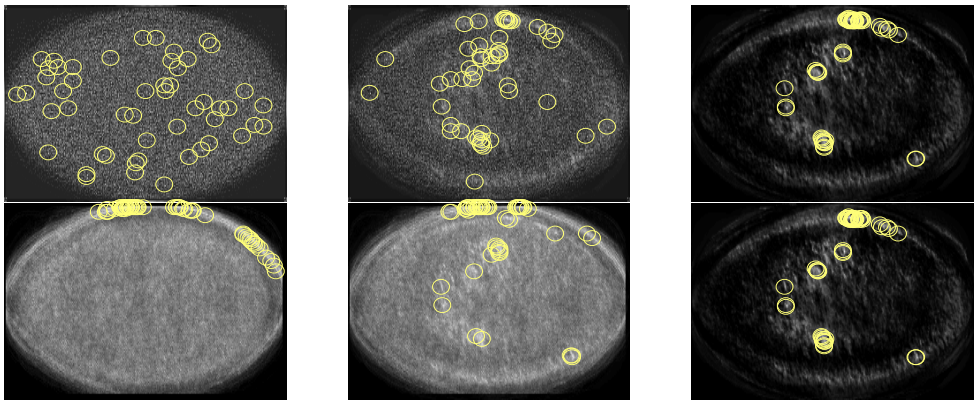
$$\begin{aligned}&\log_2|WI(q,s)|\\ =\ &\log_2(K) + (\alpha+1/2)\log_2 s - \frac{n-\alpha}{2}\log_2(1+\frac{\sigma^2}{\sigma^2 s^2}),\end{aligned}$$

where s is a scale factor, K is a constant, n is an integer, and the wavelet transformation W is contributed by the derivative of Gaussian of variance σ^2. Intuitively, the Lipschitz regularity at p is the maximum slope of $\log_2|WI(q,s)|$ as a function of $\log_2 s$ along the maxima lines converging to p. In our experiments, the scale factor s is from 2^0 to 2^4.

V. Experiments

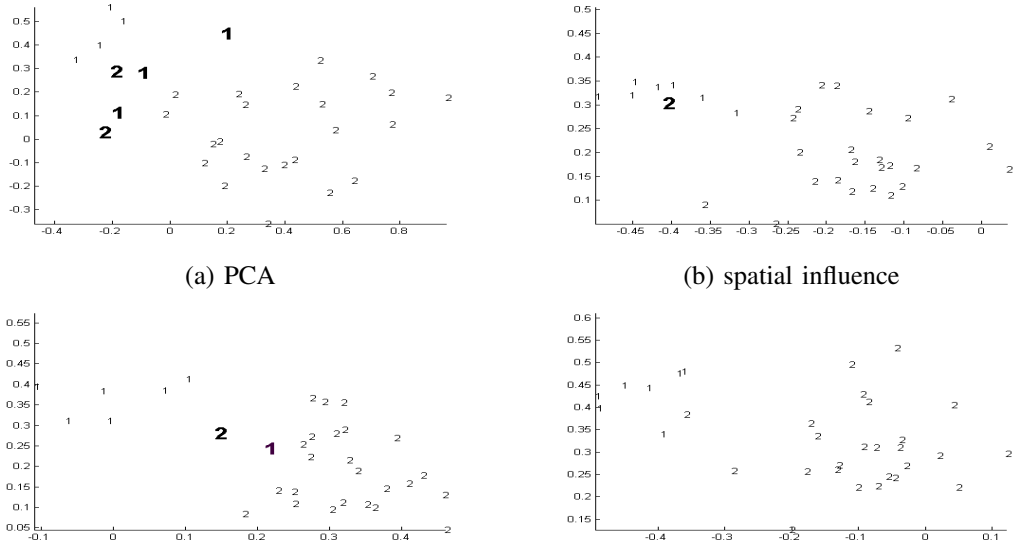
In this section, we test the proposed method by a case study of recognition of embryo stages. Our dataset has 500 images of fruit fly embryo, in three classes. The goal of classifying embryo images is to identify embryo developmental stages that is an important step towards gene expression analysis. The raw images contain severe illumination variations. We apply histogram equalization method to normalize embryo images. Recall that the nature of weak texture of embryo images motivates us to explore the opportunity of using relatively large number of features.

In our experiments, a dataset is randomly split into two: one half is used as training and validation set, and the other half as the test set. We run 5-fold on training and validation data to decide the optimal parameters: weights (bottom-up and top-down priors) and the number of feature points. To reduce the variability, the splitting is repeated 5 times and the resulting accuracies are averaged. The number of feature



(a) Spatial (b) Integrated ($\alpha = 0.5$) (c) Discriminant

Figure 1. Influence maps of embryo images overlaid by 50 best feature points. First row = gradient auto-correlation; Second row = Laplacian.



(a) PCA (b) spatial influence

(c) discriminant influence (d) integrated influence

Figure 2. Visualization of 4 different feature vectors. The bold labels indicate the data items violating linear separability. The integrated influence contributes feature vectors of best linear separability.

points (n) in our experiment is from 400 to 2000. We use nearest neighbor as the classifier.

A. *Fisher score: standard versus locality oriented*

First, we present a comparison between the standard Fisher score and the locality oriented Fisher score (l=1 or 2) in three different appearance based recognition tasks. Table I shows the results, and it is clear that the locality oriented Fisher score outperforms the standard Fisher score. We observe that the performance of norm $l = 2$ is slightly better than norm $l = 1$, in the case of pure discriminant selection. However, as we will see soon, the observation will be different when the locality oriented Fisher scores are integrated with a certain bottom-up scheme, which in turn leads to the use of both norm in the integrated model.

Furthermore, we measure the performance of locality oriented discriminant influence with different norms integrated with a certain bottom-up scheme. Fig. 3 illustrates the

discriminant methods	Embryo
standard Fisher score	0.80
LO Fisher score (l=1)	0.83
LO Fisher score (l=2)	0.84

Table I
RECOGNITION ACCURACY. A COMPARISON BETWEEN STANDARD FISHER SCORE AND LOCALITY ORIENTED FISHER SCORES.

methods	GA	Laplacian	All
selected feature	0.93(0.04)	0.91(0.05)	0.94(0.03)
feature+LDA	0.83(0.04)	0.82(0.05)	0.85(0.03)

Table II
RECOGNITION ACCURACY (WITH DEVIATION) ON EMBRYO STAGES OF INTEGRATED MODEL. THE BEST RESULT REPORTED BEFORE IS AROUND 70%.

behavior under a simple version of integrated model (spatial influence is contributed by gradient auto-correlation only). We can observe that highest accuracy is achieved by the integrated influence associated with norm $l = 1$. It is worth noting that this interesting observation occurs consistently across varied n, which reveals the benefit of introducing l in the locality oriented Fisher criterion score. In the later experiments, we use two discriminant operators, i.e., T^1 and T^2 are associated with norm 1 and 2, respectively.

B. Main results

Fig. 4 shows the validation accuracy in cross-validation, where X-axis indicates the weight α, Y-axis indicates the length of feature vectors, and Z-axis indicates the validation accuracy. Fig. 4 (a) and (b) are associated with gradient auto-correlation, and Laplacian (two spatial influence assignments), and the norm l in the discriminant influence is 2. First of all, Fig. 4 gives an example that discriminant influence does not always outperform spatial influence. More importantly, Fig. 4 shows the mutual benefit of spatial and discriminant influences, for example, the highest accuracy is always achieved by a certain degree of integration of spatial and discriminant influence. The optimal parameters for gradient auto-correlation are ($\alpha = 0.5, n = 400$), and the ones for Laplacian are ($\alpha = 0.6, n = 2000$).

Tables II shows the performance of the integrated model on three datasets, including the recognition accuracy and the deviation. The first row is associated with the vectors consisting of selected features, and the second row is associated with LDA (Linear Discriminant Analysis) representation of the feature vectors. In the second and third column, the model uses gradient auto-correlation and Laplacian, respectively. In the last column, the model uses both influence maps. The model uses the unsupervised operator, and two supervised operators. The results convince the effectiveness of the model in integrating different spatial influence maps, i.e., higher accuracy and smaller deviation. It is also worth noting that the integrated model outperforms some baseline methods. The results from Table II show LDA representation degrades the performance of the feature vectors. After all, the dimension of the LDA representation is much lower than the the dimension of the feature vectors.

VI. CONCLUSION

In this paper, we introduced a locality oriented Fisher scores via wavelet, and proposed a model to integrate spatial and discriminant influence. A motivation of our study comes from the insight that the appearances of certain objects are weakly-textured, and a few of features may not be robust with respect to the imaging variations. We explore the opportunity of using relatively large numbers of features. In experiments, we verified the effectiveness of the locality oriented Fisher scores, and showed the promise of the integration model with the tests on three different applications.

ACKNOWLEDGMENT

The work of Q. Li was supported by National Science Foundation Grant IIS-1016668 and the Summer Faculty Scholarship 10-7052 of Western Kentucky University.

REFERENCES

[1] C. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, UK, 1995.

[2] I. Daubechies. Ten lectures on wavelets. SIAM, Philadelphia, 1992.

[3] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *IEEE Computer Vision and Pattern Recognition*, pages 264–271, 2003.

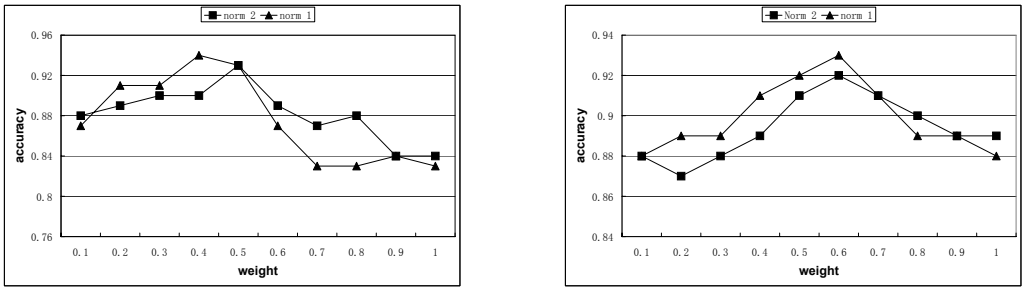
[4] T. S. Furey, N. Cristianini, N. Duffy, D. Bednarski, M. Schummer, and D. Haussler. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16(10):906–914, 2000.

[5] D. Gao and N. Vasconcelos. Discriminant saliency for visual recognition from cluttered scenes. In *Neural Information Processing Systems (NIPS))*, Electronic edition, 2004.

[6] R. Gurunathan, B. Emden, S. Panchanathan, and S. Kumar. Identifying spatially similar gene expression patterns in early stage fruit fly embryo images: binary feature versus invariant moment digital representations. *BMC Bioinformatics*, 5:202, 2004.

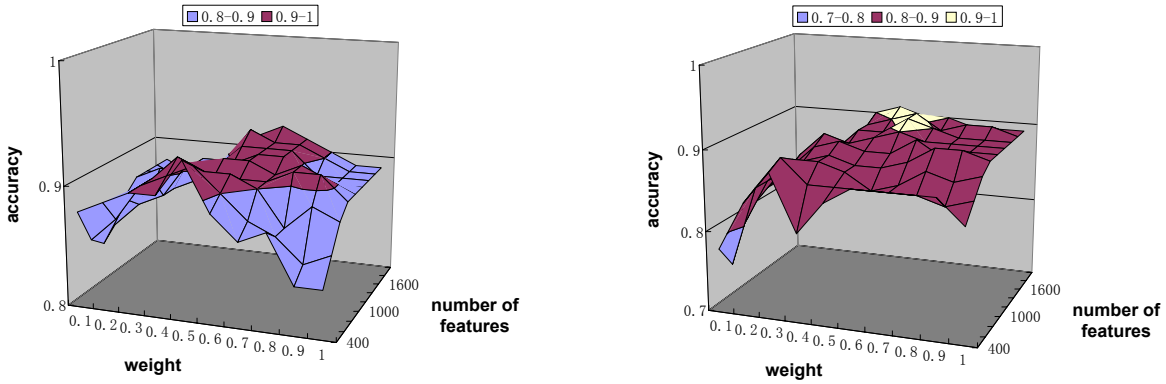
[7] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. 4th Alvey Vision Conference, Manchester*, pages 147–151, 1988.

[8] A. Holub, M. Welling, and P. Perona. Combining generative models and fisher kernels for object recognition. In *IEEE International Conference on Computer Vision*, pages 136–143, 2005.



(a) 400 features (b) 2000 features

Figure 3. A comparison between norm $l = 2$ and $l = 1$. The highest accuracy is achieved by the integrated influence associated with norm 1.



(a) Gradient auto-correlation (b) Laplacian

Figure 4. Integrated influence with norm $l = 2$ in discriminant influence. The optimal parameters for gradient auto-correlation are $\alpha = 0.5, n = 400$, and the ones for Laplacian are $\alpha = 0.6, n = 2000$.

[9] S. Kumar, K. Jayaraman, S. Panchanathan, R. Gurunathan, A. Marti-Subirana, and S. Newfeld. Best: a novel computational approach for comparing gene expression patterns from early stages of drosophila melanogaster development. *Genetics*, 162(4):2037–2047, 2002.

[10] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[11] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(7):710–732, 1992.

[12] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.

[13] K. Mikolajczyk and C. Schmid. Indexing based on scale invariant interest points. In *IEEE International Conference on Computer Vision*, volume I, pages 525–531, Vancouver, Canada, 2001.

[14] V. Navalpakkam and L. Itti. An integrated model of top-down and bottom-up attention for optimal object detection. In *Computer Vision and Pattern Recognition (1)*, pages 1–7, 2006.

[15] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of interest point detectors. *IJCV*, 37(2):151–172, 2000.

[16] N. Vasconcelos. Feature selection by maximum marginal diversity. In *Neural Information Processing Systems (NIPS)*, 2002.

[17] N. Vasconcelos and M. Vasconcelos. Scalable discriminant feature selection for image retrieval and recognition. In *Computer Vision and Pattern Recognition (2)*, pages 770–775, 2004.

[18] D. Walther, U. Rutishauser, C. Koch, and P. Perona. Selective visual attention enables learning and recognition of multiple objects in cluttered scences. *Computer Vision and Image Understanding*, 100:41–63, 2005.

[19] M. Weber. *Unsupervised Learning of Models for Object Recognition*. PhD thesis, Department of Computational and Neural Systems, Caltech, Pasadena, CA, 2000.