Top-down facilitation of multistage decisions for face recognition

Brian Parks and Terrance Boult

Preprint of paper to appear at IEEE Int. Conf on Biometrics Theory Applications and Systems (BTAS) 2010.

Abstract— **Visual processing in humans is, without a doubt, far superior that that in machines, especially when the end goal is object or face recognition. Neural results from visual object and face recognition in humans provide an excellent model for developing better techniques in machine vision. In this study, we present a particular neural result pertaining to the use of low spatial frequency (LSF) imagery to facilitate recognition of high spatial frequency (HSF) representations of faces and objects and apply it first as a general technique for the classification problem and second as a high-performance recognition method to deal with face recognition on blurry imagery. We demonstrate significant improvement over baseline results using a directly comparable published algorithm. We also discuss the problem and our technique for solving it terms of a mutually beneficial collaboration between the fields of computer vision and neuroscience.**

I. Introduction

Human beings are very good decision makers. We make decisions every day, from what we are going to eat for breakfast to what time we are going to leave work. We frequently also make decisions when we don't realize it, for instance when recognizing an object or a face. In fact, humans have to make these types of decisions so frequently that our brains have specialized to be able to make these decisions and make them well. Occasionally, we may mis-recognize an acquaintance we haven't seen in years, but on the whole, our brains are very effective at solving the face recognition problem.

In fact, most circumstances show that humans are the clear victor in the face recognition challenge over computer vision methods. [21] gives a thorough summary of many of the unconstrained face recognition scenarios in which humans perform incredibly well while machines are left in the dust. In particular, the present work will focus on one specific instance for illustration purposes: the effects of motion blur, which do not present a serious issue for humans but are crippling to machine vision systems.

It is unsurprising that blurry imagery makes the task of face recognition more difficult and thus is an interesting problem in computer vision. However, it is also unsurprising that humans seem to be considerably better at it; humans are constantly in motion and seeing at distance, yet are still able to recognize faces better than computers can in these unconstrained circumstances.

B. Parks is with the Department of Computer Science, University of Colorado at Colorado Springs, Colorado Springs, CO 80918, USA `bparks@vast.uccs.edu`

T. Boult is the El Pomar Professor of Innovation and Security at the University of Colorado at Colorado Springs and CEO/CTO of Securics, Inc., Colorado Springs, CO 80918, USA `tboult@vast.uccs.edu`

It only seems natural that a system so elegantly tuned to the problem of decision making and, in particular, object and face recognition, should become a model for implementing a similar facility in machines. Indeed, since the late 1950s, the brain's physiological structure and function formed the (albeit simple and primitive) model for the development of the perceptron [19] (described in detail in [20]) and its successor, the artificial neural network. While neural networks gained popularity, especially among the pattern recognition community [4], newer models such as the Support Vector Machine (SVM) [6] have begun to effectively replace the neural network.

This study attempts to find a balance between the decision-making model of the brain in the context of face recognition while still making use of the efficient and more practical SVM machine learning technique. We apply a result from the human visual processing community to a similar problem in machine vision, using the system of processing in the human brain as a model to conduct the recognition in multiple phases. While biometric algorithmic fusion is not a new concept in biometrics, our contribution is in demonstrating the applications and theoretical underpinnings of such recognition system in terms of the neural model presented below, demonstrating significant improvement over baseline blur results.

This study is demonstrated as follows: in Section I-A, we discuss the relevant neurological research providing inspiration for our proposed decision method and face recognition process. In Section II we discuss how this methodology can be applied to the classification problem as defined in Machine Learning and, in particular, the Support Vector Machine. We also discuss the specific problem of face recognition and, in Section III, describe experiments used to test our methodology in this domain and their results. We provide a brief discussion of the significance and future potential in Section IV.

A. Top-down Facilitation of Visual Processing in humans

On the neuroscience side, visual processing has long been a subject of study, beginning with the accidental discovery by Hubel and Wiesel in 1959 [11] of neurons sensitive to lines at varying orientations in the V1 area (primary visual cortex) of the cat. Since then, research has progressed beyond the bottom-up approach of studying the inputs to V1, the processing done in its relatively localized brain regions, and successively higher visual processing areas ultimately leading to the Medial-temporal Lobe (MTL) as a more holistic view has gained support.

In fact, an empirical study by Bar et al. [2] proposes

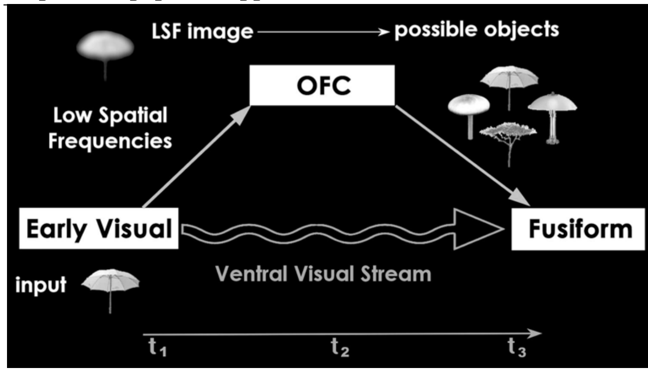


Fig. 1. Bar et. al. [1] define two pathways from optic nerve to MTL. One processes a high spatial frequency version of a visual stimulus, while the other processes low spatial frequencies 90ms prior.

an alternate pathway from optic nerve to MTL by way of the orbito-frontal cortex (OFC) and bypassing V1 that is used specifically for visual recognition tasks. This pathway, detailed further in [1], operates on low spatial frequency (LSF; what most people would describe as "blurry" or "low-resolution") images approximately 90 milliseconds before the full visual stimulus (high spatial frequency or HSF image) is processed by MTL (130 ms after stimulus onset). A more precise pathway is proposed in [12], both explaining and lending credence to the idea that the early stage of processing operates on what is essentially a low-resolution image. [17] describes this hypothesis from a higher-level perspective: OFC serves to integrate the low spatial frequency information with signals from other parts of the higher-level brain, most notably our memories, to narrow down the possible identities of the visual stimulus to a specific subset of objects with which we are familiar.

We can interpret these findings in a more precise way that is more suitable to generalization to the machine vision analogue: In machine vision problems, the role of "memory" is taken over by a machine learning classification model that is trained on samples of images much the way a human observes faces over the course of his or her lifetime. The LSF image is used to choose a set of candidate identities for an unknown input image from higher metacognitive memory in the OFC that are used to seed the remainder of the recognition process and arrive at a single identity. In this way, a limited amount of data can be used to quickly reduce the number of possible identities prior to a final recognition phase using the full amount of data.

II. Model in terms of Machine Learning

Recently, conclusions from neuroscience have pervaded, not only the way in which features in a machine learning problem are classified, but also the way in which features to be classified are generated. Computer vision has begun to use spectral analyses of the primary visual cortex [8] in the context of object and face recognition problems [16] to generate particularly discriminative feature vectors for use with Support Vector Machines. We can extend this model in terms of the system described above.

If we look deeper into the neural mechanisms that allow for this heightened ability to perform face recognition, as reviewed in section I-A, we see that they are not as simple as most face classification problems make them out to be. In particular, we see that the decision is made in multiple stages. First, raw data that has undergone a minimum of preprocessing is used to prime the decision mechanisms by reducing the decision space. Enough information is obtained from this rough approximation to determine that a human face is not that of a dog and a face with facial hair is that of a mature male.

Given this reduced decision space, the processed information can then be used to distinguish between the individual classes that have not been discarded by the rough decision phase. This approach can further be generalized to allow for multiple stages of processing and even (as may be the case in the human brain) to use different classification methods trained on different sets of features for each stage in the decision. This study demonstrates the latter approach, using one model to classify LSF images and a second to classify the HSF original image.

A. Support Vector Machines

As proposed in [6], the Support Vector Machine is a mathematical model of machine learning that operates by learning a function to map feature vectors of dimension n into an m-dimensional space, where $m > n$, and a hyperplane in the m-dimensional space to separate examples of a positive class C^+ from those of a negative class C^-. A classifier thus generated can then be generalized to solve a k-class problem by taking k classifiers, each machine M_i of which decides the class C_i. This is referenced in the Machine Learning literature as *one-versus-rest*.

During the training phase, all component SVMs in the multiclass SVM are trained simultaneously, depending on whether or not the specific piece of training data (we will use the term *gallery vector* for the remainder of the paper) represents a positive instance of that class or not. Thus, during classification of test data (*probe vector*), the immediate result is an array D, of size k, of *marginal distances*, each d_i of which represents the distance from the probe vector to the region in m-dimensional space representing class C_i. As such, lower numbers indicate increased certainty that the probe is a member of class C_i and the class corresponding to the lowest marginal distance is the ultimate output of the classification. By negating these marginal distances, it is trivial to show the equivalence of these distances to scores expected in the general machine learning classification problem.

Further details on the definition, structure, and operation of support vector machines are documented in [7].

B. Problem Definition

In a traditional face recognition scenario, the problem may be defined as follows: given a set of subjects (classes) C and a gallery G such that, for for every $C_i \in C$ there exists a $g_j \in G$, where g_j is a member of class C_i, classify an unknown probe image, p, as a member of one and only one

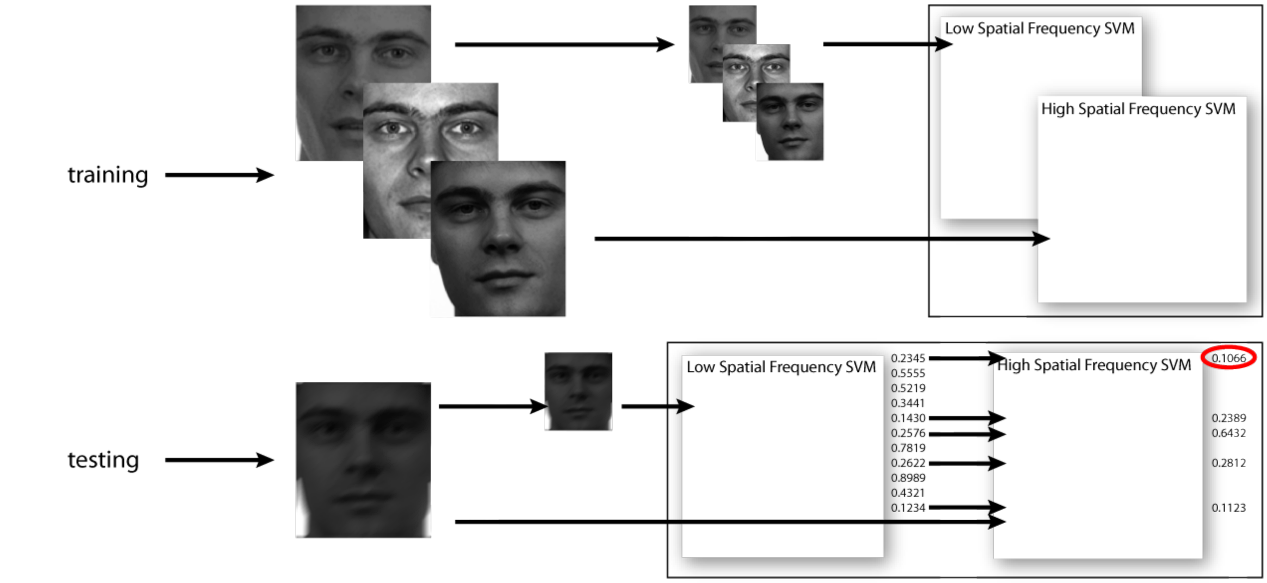


Fig. 2. Flowchart of our Machine Learning implementation of top-down facilitation. The training process begins with several input images of each subject and trains two SVMs – one on a downscaled version of the image and one on the full-sized image. The full model (rectangle at right) is then used for testing as demonstrated by the rectangle on the bottom. A downscaled image is classified using the LSF SVM and the marginal distances are used to seed the classification process of the HSF SVM on the original input image.

class in C. Usually, this is done by mapping each g_j and p, all of which are of the same type of datum of arbitrary, but consistent, dimension, to a feacture vector $\mathbf{v}$ of arbitrary, but consistent, dimension. The problem is then reduced to learning a function of the form:

$$f(\mathbf{v}) : \mathbf{v} \to C_i \in C \tag{1}$$

However, this scenario can be generalized to allow for multistage decisions. In order to perform this operation, we modify Equation 1 slightly:

$$f^{(n)}(\mathbf{v}^{(n)}, C^{(n)}) : \mathbf{v}^{(n)} \to C^{(n+1)} \subset C^{(n)} \tag{2}$$

In Equation 2, $C^{(n)}$ represents a subset of C, where $C^{(0)} = C$ and $C^{(N)}$ contains only one element, the result of classification. $\mathbf{v}^{(n)}$ represents a different feature vector for each stage of the decision, presumably representing increased resolution or preprocessing (ultimately represnet-ing a larger amount of data per input image), similar to the proposed progression in the human visual recognition pathways. This function, once learned, is then solved for all values of $n \in \{0, \ldots, N-1\}$.

In the specific case of blurred imagery, as demonstrated here, we choose $N = 2$ and define $\mathbf{v}^{(0)}$ to be the features of an image scaled down to one-sixth of its original size (and, thus, both digital and optical resolution representing $\frac{1}{36}$ the original amount of data) and $\mathbf{v}^{(1)}$ to be the features of the image either as provided to the recognition core (with some form of motion blur) or after it has been preprocessed (to compensate for or remove motion blur).

The underlying model in our experimental setup is a multiclass Support Vector Machine both because it is easy to separate into parts by component class and because of its current popularity. In order to implement the multi-stage decision component, we actually train two multiclass SVMs. The first is trained on LSF imagery generateed by downsampling each of the original gallery images. This will provide a rough set of potential classes during the classification phase. It is worthwhile to note that, despite the counterintuitive nature of using lower-resolution imagery to refine the recognition task and boost overall recognition accuracy, Beveridge et al. [3] have demonstrated that low-quality (out-of-focus) imagery may actually lead to better recognition rates in certain situations.

The second SVM is trained on the raw gallery images at full size. The component SVMs thus trained are then selected based on the marginal distances generated by classifying an LSF version of a probe image with the first SVM. These selected SVMs are assembled into a new multiclass SVM with a significantly reduced decision space. To choose the members of $C^{(1)}$ (represented by this re-assembled SVM), we select the classifiers from the SVM modeling $f^{(0)}$ that produce the top k scores, varying k between sets of experiments.

We choose values of k with the intent of excluding outliers that could potentially confuse the second classification as a byproduct of the blurred original imagery. Thus, we have chosen values of 10, 30, 60, and 100. Any more than 100 classes does not provide a significant reduction in the dimension of the decision space, while any fewer than 10 does not retain enough of the candidate classes to be meaningful (LSF images of objects, especially of faces, tend to look the same across a significant cross-section of classes, but also provide enough variation from an even more significant portion of incorrect classes to provide meaningful top-down information). The chosen values of k do not represent consistent spacing along a range because results at other values do not represent a significant deviation from the trends shown by these choices of values.

Fig. 3. A member of the original FERET set, prior to preprocessing and blurring

Once a reduced set of potential classes is obtained, we construct a feature vector either from the original image or from an image obtained as a result of deblurring using techniques such as those described in [13], [14], [9], which is then used as input to an SVM constructed of those component SVMs (modeling $f^{(1)}$) correspondng to classes in $C^{(1)}$ to determine the final classification. This pipeline is shown graphically in figure 2.

III. Experiments and Results

The program used as a framework for the feature generation and classification is a fork of the Python code published with [16] using Gabor responses as features for a multiclass SVM. The underlying SVM implementation is a SWIG interface from python to the libsvm library encapsulated in the Python library PyML.

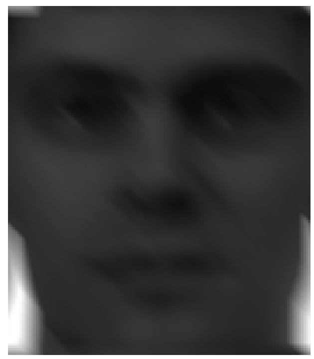
We demonstrate results for all stages of the current study on two well-known public datasets. The FERET (Face Recognition Technology) [15] set was chosen due to its relatively constrained nature; all images used were frontal and under fairly consistent lighting conditions. This allows us to demonstrate the substantial effect of blur on recognition rates while also providing a significant margin within which to demonstrate improvement as a result of our methodology.

Due to the constraints of our SVM-based classification method, a gallery of more than a single image was required. To reduce the potential for an outlier to have potentially disastrous effects on the training of the SVM while still maintaining a relatively small gallery size, we used three gallery images per subject. As such, it was necessary to deviate from the published protocol for FERET.

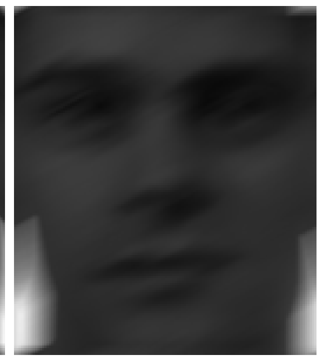
The following protocol was designed and used for testing: subjects for whom the data set contained fewer than four

(a) 5 pixels

(b) 10 pixels



(c) 15 pixels



(d) 20 pixels

Fig. 4. A member of the FERET set under various motion blur extents. Note that the angle of blur was randomized during the processing phase.

images were discarded. For each of the remaining subjects, a set of four images were chosen by an alphabetic sort on the names given in the original data set. Of these four images, the first three comprised a subject's gallery; the last was used as a probe image. Thise subset has been dubbed FERET240.

Note that this protocol deviates slightly from the FERET protocol in that the standard FERET "fa" subset is generally intended for gallery, while our protocol generally used "fa," "fb," and "fc" as gallery and some of the more difficult FERET images as probes. This breakdown is merely coincidence; the alphabetic sort as described above make the distinction between probe and gallery largely arbitrary.

All images in the resulting dataset were geometrically normalized, prior to processing, using the CSU Face Identification Evaluation System [5]. Subsequently, the "probe" subset of the FERET240 set was processed using MATLAB's `fspecial` function to simulate motion blur at various specific blur lengths and random angles to create four new datasets corresponding to 5, 10, 15, and 20 pixel blur.

To establish baseline results on the datasets used in this study, a version of the code base published with [16] was modified only to allow control over which images were used as gallery and probe. The results of the otherwise-unchanged recognition core on our FERET240 dataset are summarized in Table I. As shown in this table, blur greater than 10 pixels is sufficient to reduce baseline recognition by a statistically significant percentage. Figure 4 shows an image from the FERET set at various blurs.

Blur	None	5 pixels	10 pixels	15 pixels	20 pixels
Rank 1	97.50	95.00	75.00	39.58	16.67

TABLE I

BASELINE RECOGNITION RESULTS ON ALL VARATIONS OF FERET240

k	None	05px	10px	15px	20px
10	92.00	95.42	87.08	71.67	57.08
30	96.00	95.42	84.58	62.08	42.50
60	96.00	95.42	82.92	53.33	34.58
100	96.00	94.58	82.08	49.17	31.25

TABLE II

RECOGNITION RESULTS ON FERET240 USING THE PROPOSED SYSTEM USING DIFFERENT VALUES OF k

A. Multiphase recognition

We have implemented the process described in Section II-B in a high-level protoyping language, based on the code base provided with [16] and following the experimental protocol described in Section III for each of the four values of k noted in Section II-B.

During the recognition phase of each experiment, each probe image was first classified using its one-sixth resolution version, giving a set of marginal distances corresponding to the distance in multi-dimensional space from the blurred probe's feature vector and each component SVM. The class labels from the SVMs most likely to correspond to the correct class (as determined by the appropriate threshold) were used to choose the constituent SVMs from the second multiclass SVM, which then classified the original input image. The result of this classification was used as the final result of classification, regardless of the scores generated from the initial classification.

Table II summarizes the results on the FERET240 set. It is interesting to note that the classes whose models do not provide a score in the top 10 during the first classification phase seem only to confuse the second classification phase, which suggests that reducing the digital and optical resolution of a degraded image may provide a more accurate classification than attempting to compare a degraded probe against clean gallery images. While these results represent marked improvement over the baseline blurred recognition results, other attempts to solve the blur problem such as [10] demonstrates an equivalent or better performance increase on the same data via the simple procedure of deblurring and recognizing using a single unmodified SVM.

However, this discrepancy can be made up by deblurring the original input probe using a Wiener deconvolution with a point spread function generated from the blur parameters as described in [10]. Since this study is not on deblurring methods, we simply use the ground truth blur parameters obtained when generating the datasets. The rank 1 recognition results using deblurring with both a single-phase SVM (1PS) and our multi-phase SVM (MSVM; $k = 30$) are summarized in Table III. We choose $k = 30$ because retaining 30 classes provided better results than retaining only 10 classes for the second phase. This is due to the increase in similarity of a deblurred image to its respective gallery images, which decreases the extent to which it

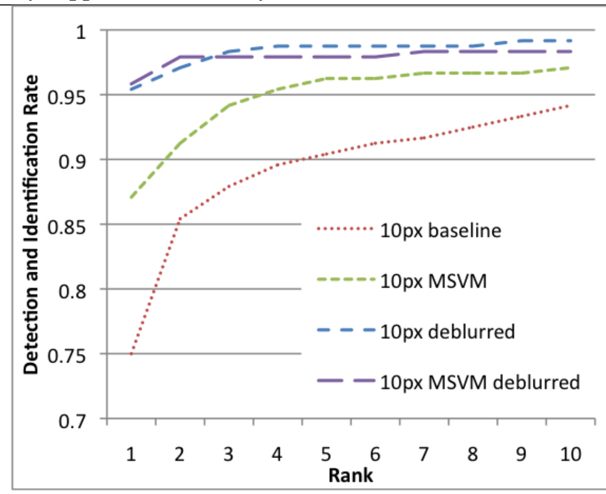


Fig. 5. Cumulative Match Curve for the 10-pixel blur case. This demonstrates how the method described in this paper forces the vast majority of its matches closer to Rank 1, whereas use of a single SVM creates a smoother curve, losing accuracy fairly consistently esp. in ranks 4 through 10. Note that the y-axis ranges from .7 to 1.0.

	05px	10px	15px	20px
MSVM	95.42	95.83	95.42	94.17
1PS	97.50	95.83	94.58	92.50

TABLE III

RECOGNITION RESULTS ON USING DEBLURRED IMAGES AS INPUT TO A STANDARD MULTI-CLASS SVM (1PS) AND TO THE SECOND PHASE OF A MULTI-PHASE SVM (MSVM)

exhibits similarity to incorrect classes.

This means that we can loosen the restrictions on the number of classes we use to seed the second phase without having to worry about introducing classes that will only serve to confuse the recognition process. In fact, it allows the correct classes that do not fall within the top 10 scores (but do fall within the top 30 scores) to be counted and used to seed the second phase of classification. Thus, the increase in recognition between the case with no deblurring and the case with deblurring is due to both the deblurring and the reduction in the number of correct classes accidentally eliminated in an effort to reduce confusion in the second phase.

The results of all four recognition methods presented in this paper are shown in Figure 5 for a single blur length. This shows how the use of the multi-phase support vector machine forces its matches closer to rank 1, whereas a single SVM exhibits a smoother curve, but suffers from loss of overall accuracy due to the fairly consistent slope especially between ranks 4 and 10.

IV. CONCLUSION

In this study, we have demonstrated that it is possible to model a machine vision system for the face recognition problem after a simplified model of multiple brain pathways through a form of biometric algorithmic fusion. We have shown that without performing computationally-intensive calculations to determine blur parameters from an image, it is still possible to achieve substantial increases over baseline recognition scores on blurred imagery. We have also demonstrated that it is possible to reclaim even more of the lost performance by also performing deblurring, demonstrating that these results exceed those obtained from deblurring and using a single recognition model.

While the method proposed here bears certain similarities to so-called ”multi-pass” recognition methods, academic literature on the subject is surprisingly slim. It is apparent that ”multi-pass” techniques are being used in some of the leading commercial face recognition algorithms, yet formal discussion of their benefits, strengths, weaknesses, and past and present development is limited. This study attempts to bring these concepts into the realm of formal academic research by showing their high applicability to, at the very least, a specific class of unconstrained face recognition problem.

This work also provides a potential alternative to modifying a large set of gallery images based on certain properties of each input probe image. For example, [18] demonstrates the drastic effect of inaccurate eye localization in face recognition scenarios. One way to approach this is to perturb the eye coordinates in the gallery, which is time-consuming and expensive. Using the system proposed in this study, however, the eye localization errors would be minimized or eliminated on an LSF version of an input image, providing a similar effect on recognition rates as those on motion blur studied here. We leave specific experimentation in this situation to future work.

Future work will also entail exploring other methods for using the result of a prior classification to seed a later one that may more closely reflect continuing research into the neural underpinnings of top-down facilitation. In addition, the technique proposed in this study may find applications in other unconstrained machine vision problems, such as pose, expression, changes in faces over time, and other such conditions that would not heavily impact a low spatial frequency image.

Perhaps more importantly, we have shown that, while the underlying components used in a scenario such as the motion blur problem may not necessarily reflect the neural systems they simulate, they can be combined in a way that does reflect the higher-level findings from the field of neuroscience. Brain-inspired models of computer vision are, like all models, simplifications of the real thing. Due to the extreme complexity of the brain, they are even more oversimplified than the average model. For instance, there is considerably more processing between retina and primary visual cortex than a Gabor filter can describe and there is even more processing after that stage before the signals reach MTL and can be thought to represent “recognition.” However, the hope is that, as we increase the extent to which our models of machine vision mirror the processes involved in recognition in the human brain, the accuracy of recognition in unconstrained environments will increase drastically. This study represents another step in the progress of the field of computer vision as it recognizes the benefits of close collaboration with the fields of both computational and biological neuroscience.

ACKNOWLEDGMENTS

This work was supported by ONR MURI N00014-08-1-0638, ONR STTR N00014-07-M-0421, and NSF PFI #0650251. We would also like to thank Walter Scheirer and Brian Heflin for their help in preparing and writing various parts of the code base and Jugal Kalita, who provided valuable feedback on an earlier draft of this work.

REFERENCES

[1] M. Bar, K. M. Kassam, A. S. Ghuman, J. Boshyan, A. M. Schmid, A. M. Dale, M. S. Hamalainen, K. Marinkovic, D. L. Schacter, B. R. Rosen, and E. Halgren. Top-down facilitation of visual recognition. *Proc. of the National Academy of Sciences*, 103(2):449–454, 2006.

[2] Moshe Bar. A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, 15(4):600–609, 2003.

[3] J. R. Beveridge, G. Givens, P. J. Phillips, and B. Draper. Focus on quality, predicting FRVT 2006 performance. In *Intl. Conf. on Automatic Face and Gesture Recognition*, 2008.

[4] Christopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.

[5] D. Bolme, R. Beveridge, M. Teixeira, and B. Draper. The CSU Face Identification Evaluation System: Its purpose, features and structure. In *ICVS*, pages 304–311, 2003.

[6] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[7] Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000.

[8] John G. Daugman. Two-dimensional spectral analysis of cortical field receptive profiles. *Vision Research*, 20(10):847–856, 1980.

[9] W. Fawwaz A M, T. Shimahashi, M. Matsubara, and S. Sugimoto. PSF estimation and image restoration for noiseless motion blurred images. In *SSIP*, pages 1–7, 2008.

[10] Brian Heflin, Brian Parks, Walter Scheirer, and Terrance Boult. Single image deblurring for a real-time face recognition system. In *36th IEEE Industrial Electronics Society Conf. (IECON-2010)*, 2010.

[11] David Hubel and Torsten Wiesel. Receptive fields of single neurones in the cat’s striate cortex. *Jour. of Physiology*, (148):574–591, 1959.

[12] Kestutis Kveraga, Jasmine Boshyan, and Moshe Bar. Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, 27(48):13232–13240, 2007.

[13] Anat Levin, Yair Weiss, Fredo Durand, and William T. Freeman. Understanding and evaluating blind deconvolution algorithms. *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2009.

[14] R. Lokhande, K.V. Arya, and P. Gupta. Identification of parameters and restoration of motion blurred images. In *ACM. Symp. on Applied Computing*, pages 301–315, 2006.

[15] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi. The FERET evaluation methodology for face recognition algorithms. *IEEE TPAMI*, 22(10), 2000.

[16] Nicolas Pinto, James J. DiCarlo, and David D. Cox. How far can you get on a modern face recognition test set using only simple features? *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2009.

[17] Rasha Abdel Rahman and Werner Sommer. Seeing what we know and understand: How knowledge shapes perception. *Psychonomic Bulletin and Review*, 15(6):1055–1063, 2008.

[18] Terry Riopka and Terrance Boult. The eyes have it. In *WBMA ’03: Proc. of the 2003 ACM SIGMM workshop on Biometrics methods and applications*, pages 9–16, New York, NY, USA, 2003. ACM.

[19] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386–408, 1958.

[20] Frank. Rosenblatt. *Principles of neurodynamics; perceptrons and the theory of brain mechanisms*. Spartan Books Washington, 1962.

[21] P. Sinha, B. J. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: 19 results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, 2006.