Hindawi Publishing Corporation
EURASIP Journal on Applied Signal Processing
Volume 2006, Article ID 31062, Pages 1–11
DOI 10.1155/ASP/2006/31062

Adaptive Markov Random Fields for Example-Based Super-resolution of Faces

Todd A. Stephenson[1,2] and Tsuhan Chen[1]

[1] *Electrical & Computer Engineering Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890, USA*
[2] *ReallaeR, LLC, P.O. Box 549, Port Republic, 20676 MD, USA*

Received 21 December 2004; Revised 1 April 2005; Accepted 5 April 2005

Image enhancement of low-resolution images can be done through methods such as interpolation, super-resolution using multiple video frames, and example-based super-resolution. Example-based super-resolution, in particular, is suited to images that have a strong prior (for those frameworks that work on only a single image, it is more like image restoration than traditional, multiframe super-resolution). For example, hallucination and Markov random field (MRF) methods use examples drawn from the same domain as the image being enhanced to determine what the missing high-frequency information is likely to be. We propose to use even stronger prior information by extending MRF-based super-resolution to use adaptive observation and transition functions, that is, to make these functions region-dependent. We show with face images how we can adapt the modeling for each image patch so as to improve the resolution.

1. INTRODUCTION

Early work on enhancing low-resolution images addressed increasing the resolution of the image without any specific outside information related to the image domain. Methods such as linear interpolation [1] first reproduce the existing pixels to produce a magnified image and then smooth the new image.

In increasing the resolution of a video frame, however, outside information is available. That is, its neighboring frames typically contain slightly different information that can be used to increase the resolution of the center frame [2]. In contrast to interpolation, this method actually adds information that was lost when the image was taken. This approach is also appropriate when we have neighboring cameras instead of neighboring video frames recording the same scene. The work in [3] expanded multiframe super-resolution, in part, by using a Huber-Markov random field (HMRF) to define a simple prior distribution that gives low probabilities for high frequencies.

While multiple video frames may not always be available, multiple related images from the same domain may be of use instead. Example-based super-resolution [4] uses the known characteristics of this domain (i.e., the prior distribution) to perform specialized enhancement. They learn the priors from a database of high-resolution images from the same domain (this is in contrast to priors defined by hand [3]). Statistical pattern recognition methods are then used for example-based super-resolution.

Markov random fields (MRFs) [5] are one tool for example-based super-resolution. By dividing a new low-resolution image, and the unknown high frequency counterpart each into corresponding patches, two functions can be defined: the observation function ϕ and the transition function ψ. The observation function gives a score for how well a candidate high-frequency patch matches the known low-resolution patch while the transition function gives a score for how well a candidate high-frequency patch matches a candidate high-frequency patch of a neighbor. Belief propagation [6] on the MRF produces the most likely high-frequency patch to associate with each known low-resolution patch such that neighboring patches are "compatible" with each other. As the MRF only acts on a single image, this type of example-based super-resolution is not a traditional, multi-image super-resolution algorithm but, rather, a form of image restoration.

Hallucinations [7] can also be used for example-based super-resolution. As enhancement of the faces takes advantage of the images being cropped, a low-resolution face is enhanced using the database face that is the closest to it.

Then, the high-frequency components of that closest face are used to enhance the given face; as multiple images are assumed to be available, the multiple frame super-resolution of [3] is also used. In contrast to [4], this method uses deterministic methods to infer the high-frequency components of a low-resolution image. Combining ideas from [4, 7], [8] assigned a different set of candidate patches for each low-resolution patch in the MRF.

The main contribution of this paper is in adapting ϕ and ψ to be region-dependent in the cropped face images. Instead of using the standard method of having a single global observation function ϕ and a single global transition function ψ, we show how to adapt them for each patch in the face. This differs from [4] in that there is a strong prior for each respective patch in the MRF. This differs from [8], first, in adapting ψ and, second, in pooling together the candidate patches for each ϕ from similar locations (where "similar" can be defined by the distance in the spatial domain or in the pixel/feature domain); this makes ϕ region-dependent instead of just location-dependent (where location in this sense refers to a single patch). Also, this differs from [7] in that we are doing a sort of local hallucination: traditional hallucination enhances the whole face using information from only one face in the database, but here we let each local patch adapt itself using a different face in the database.

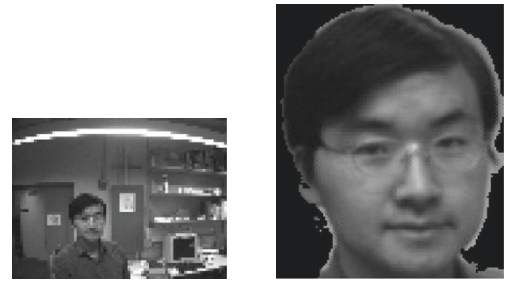
As MRFs are a type of graphical model (GM) [9], we have at our disposal, for current and future investigations, the wide variety of GM and machine learning algorithms that have been presented in the literature. For example, we can adapt ϕ by clustering certain patches together using either hand-labeling or automated clustering techniques, such as K-means clustering. The K clusters indicate the K (noncontinuous) regions of the face that are most alike in their pixel values. Patches in the same region can be jointly adapted to handle the features specific to that region. Also, we can adapt ψ using, for example, information-theoretic criteria to determine which areas of a face are compatible. The patch pairs with high mutual information can be put in the same neighborhood, even if they are in different areas of the face image.

In this paper, we describe the super-resolution problem in Section 2 before presenting how our adaptive MRFs address this problem in Section 3. In Section 4, we show the results of using these adaptive MRFs to enhance low-resolution faces. We conclude in Section 5.

2. SUPER-RESOLUTION

2.1. Preprocessing

In many domains, such as that of surveillance video, we need to extract and enhance a small object, such as a face, from a low-resolution frame (see Figure 1). As object detection [10], specifically face detection, is beyond the scope of this work, we assume that the face has been extracted and cropped. While there are different techniques available for super-resolution as outlined earlier, we summarize our baseline framework as used elsewhere [4]. Let



(a) (b)

Figure 1: Illustration of super-resolution of faces in a low-resolution video frame. (a) Low-resolution frame. (b) Face extracted, enlarged, and enhanced (simulation).

$S = \{G_0^1, \ldots, G_0^n, \ldots, G_0^N\}$ be the database (prior distribution) of N high-resolution images, with G_0^n an arbitrary image in the 0th level (the highest resolution level) of the Gaussian pyramid for image n. For the MRFs, we need the normalized high-frequency information $\widehat{H_P^n}$ and the normalized mid-frequency information $\widehat{M_P^n}$ for level P of the Gaussian pyramid that the input image occurs at. We generate them as follows.

(1) Blur and downsample G_0^n, by a factor of 2^P in each dimension, to obtain G_P^n. G_P^n is then upsampled using bilinear interpolation to obtain $G_P^{n\uparrow}$, which is the same size as G_0^n. This can then be used to determine the lost high-frequency information H_P^n in the pixel domain:

$$H_P^n = G_0^n - G_P^{n\uparrow}. \tag{1}$$

It is the task of super-resolution to recover H_P^n.

(2) High-pass filter $G_P^{n\uparrow}$. As it is assumed that the low-frequency information L_P^n of $G_P^{n\uparrow}$ is not needed to recover H_P^n from step (1), $G_P^{n\uparrow}$ is high-pass filtered to obtain the mid-frequency information M_P^n; that is, M_P^n is a band-pass filtered version of G_0^n (see Figure 2). Thus, H_P^n will be inferred using only M_P^n:

$$P(H_P^n \mid M_P^n, L_P^n) = P(H_P^n \mid M_P^n). \tag{2}$$

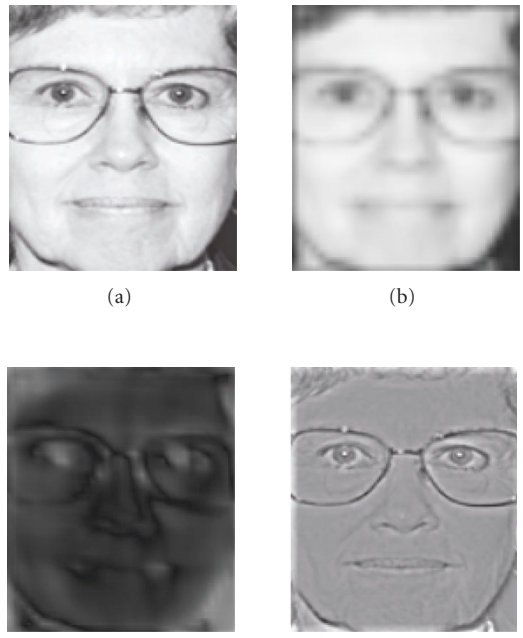
(3) Normalize the contrast in M_P^n and H_P^n. As it is assumed that the image contrast in the known M_P^n does not help to predict the unknown H_P^n, we normalize their contrast using $E(M_P^n)$, the blurred energy information of M_P^n:

$$\widehat{H_P^n} = \frac{H_P^n}{E(M_P^n)}, \tag{3}$$

$$\widehat{M_P^n} = \frac{M_P^n}{E(M_P^n)}, \tag{4}$$

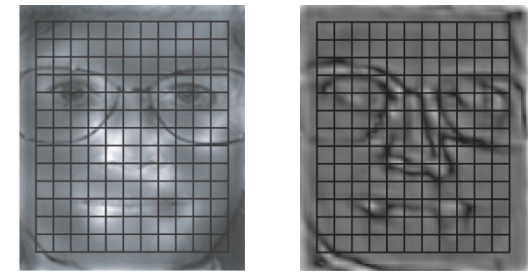
$$E(M_P^n) = (M_P^n)^2 * F. \tag{5}$$

$E(M_P^n)$ is formed by squaring the pixels of M_P^n (indicated by $(M_P^n)^2$) and then by applying a 15×15 blurring filter F.



(a) (b) (c) (d)

Figure 2: (a) The high-resolution face image G_0^n, (b) the low-resolution face image $G_2^{n\uparrow}$, (c) the mid-frequency face image M_2^n, and (d) the high-frequency face image H_2^n. The goal is to infer (reconstruct) the missing high-frequency components in (d). Image (d) has been normalized so that pixel differences of 0 have a pixel value of 128.



(a) (b)

Figure 3: (a) $\widehat{H_2^n}$ patches. (b) $\widehat{M_2^n}$ patches. Patches used in this work: 11×11 pixel patches were used for the high-frequency images, with one pixel overlap, while 13×13 pixel patches were used for the mid-frequency images, with a three pixels overlap. Corresponding high- and mid-frequency patches had the same center pixel. For simplicity, the above figure is plotted with 10×10 pixel patches, as there is a shift of 10 pixels between bordering patches. Also, to avoid artifacts from the downsampling process, no patches were placed near the border of the images.

While the above is used for preprocessing the training images, it is also used for testing the MRFs with image $\star$. That is, $\widehat{M_P^\star}$ is used as the MRF's observations; $\widehat{H_P^\star}$ is withheld from the belief propagation and is used only to evaluate the inferred results of the MRF.

2.2. Enhancement

Super-resolution of $\widehat{M_P^\star}$, where image $\star$ is an image not in S, is performed on local patches of the images, as indicated in Figure 3. The unknown target $\widehat{H_P^\star}$ is divided into 11×11 pixel patches, denoting $\widehat{H_P^\star}[i]$ for an arbitrary patch i. For each target patch i in $\widehat{H_P^\star}$ to infer, a 13×13 pixel patch $\widehat{M_P^\star}[i]$ is taken from $\widehat{M_P^\star}$ such that the center pixels of $\widehat{H_P^\star}[i]$ and $\widehat{M_P^\star}[i]$ have the same coordinates. As super-resolution in this work is probabilistic, the observation function ϕ is determined using a distribution over the training samples S. Note that in the baseline system every patch i uses the same ϕ, regardless of the location (or region) of i in the face image. As shown in [7, 11], if S is from a different domain than the image being enhanced, then the image may be enhanced incorrectly. As the observation and transition functions in our work are not strict probabilities (their summation does not equal one), we avoid the use of the word "distribution" below.

One of the functions used in this framework is the distance between the known patch $\widehat{M_P^\star}[i]$ and each high-frequency patch candidate $\widehat{H_P^{n'}}[i']$ from patch i' of image n' in S:

$$d_O\left(\widehat{M_P^\star}[i], \widehat{H_P^{n'}}[i']\right) = \left\|\widehat{M_P^\star}[i] - \widehat{M_P^{n'}}[i']\right\|. \tag{6}$$

So, to determine this distance, we compute the distance between $\widehat{M_P^\star}[i]$ and the vectorized version of $\widehat{M_P^{n'}}[i']$ (not the candidate $\widehat{H_P^{n'}}[i']$).

For each patch i, the high-frequency patch $\widehat{H_P^{n'}}[i']$ with the smallest distance can then be used to reconstruct the high-resolution image.

(1) Join all of the selected high-frequency patches into a single high-frequency image $\widehat{H_P^\star}$.

(2) Add the original contrast by multiplying $\widehat{H_P^\star}$ pixelwise by $E(M_P^\star)$, the contrast normalization matrix, to obtain the estimated $\overline{H_P^\star}$.

(3) Add the inferred high-frequency patch $\overline{H_P^\star}$ to the low-resolution $G_P^{\star\uparrow}$ to obtain the estimate $\bar{G}_0^\star$:

$$\bar{G}_0^\star = G_P^{\star\uparrow} + \overline{H_P^\star}. \tag{7}$$

3. ADAPTIVE MARKOV RANDOM FIELDS

3.1. Markov random fields

The algorithm outlined in Section 2.2 is actually incomplete as it does not take into account the relation between neighboring high-frequency patches. What is needed is to use a model which attempts to smooth neighboring patches using ψ and, hence, better model all high-frequency patches. In other words, we use a Markov random field (MRF) [5]; see Table 1. In doing so, we want to have patches in the unknown $\widehat{H_P^\star}$ to overlap by one pixel for modeling (9) below. With an MRF, we are concerned with modeling two different

Table 1: Benefit of transition function ψ: super-resolution of 38 × 33 pixel images to 150 × 130 pixels (level G_2 to level G_0) showing mean-squared error (MSE) for the whole image. Results are given using bilinear interpolation, using only the observation function ϕ, and using a standard MRF [4]. The percent reduction ("Red.") is with respect to bilinear interpolation. Results are from all 100 images in our test set.

Enhancement method	MSE	Red.
Bilinear interpolation	58.9	—
Observation only (ϕ)	64.6	−9.7%
Baseline MRF (ϕ & ψ)	54.3	7.7%

things with respect to each patch i: the observation ϕ, based on (6), and the transition ψ:

$$\phi(i = i') = \exp\left(\frac{\left[d_O\left(\widehat{M_P^\star}[i], \widehat{H_P^{n'}}[i']\right)\right]^2}{\sigma_{O_i}}\right); \quad (8)$$

$$\psi(i = i', j = j'') = \exp\left(\frac{\left[d^*\left(\widehat{H_P^{n'}}[i'], \widehat{H_P^{n''}}[j'']\right)\right]^2}{\sigma_{Ti}}\right). \quad (9)$$

We model this transition between two patch candidates: $\widehat{H_P^{n'}}[i']$ from training image n' and $\widehat{H_P^{n''}}[j'']$ from training image n''. σ_{O_i} is chosen based on the distances between $\widehat{M_P^\star}[i]$ and the closest patches to it in S; $d^*(\cdot)$ indicates the distance only between the pixels in the overlap region; and σ_{Ti} is chosen so that 10% of the possible transitions for i will have $\psi(i', j'') > 0.1$. In our baseline system, we define $\mathrm{N}(i)$, the neighborhood of i, as the four patches bordering i to its left, right, top, and bottom. In two of our proposed systems, we expand this definition to include long-distance "neighbors" either defined by hand or learned using information theoretic criteria.

As exact inference in an MRF is computationally infeasible, approximation methods are generally used [12]. Approximate probabilistic inference in the MRF is achieved by each patch i passing "messages" $m(i, j = j')$ to each of its neighbors for each value j' of each neighbor j:

$$m(i, j = j') = \sum_{i'' \in C_i} \phi(i = i'')\psi(i = i'', j = j') \cdot \prod_{k \in \mathrm{N}(i) \setminus j} m(k, i = i''), \quad (10)$$

where C_i indicates the top N closest candidate patches from S of patch i (in this work, $N = 20$). The "loopy-propagation" algorithm of [4, 11] proceeds iteratively, first, by each patch i simultaneously sending off messages $m(i, j = j')$ to each neighbor j and for each possible value j' and, second, by each patch i receiving those messages (e.g., $m(j, i = i')$) just sent to it and updating its belief in its own patches. The messages entering patch i from each of its neighbors are used to calculate the belief (i.e., the probability) of i's high-frequency patches given each neighbor j (hence the term "belief propagation"):

$$b(i = i') = \prod_{j \in \mathrm{N}(i)} m(j, i = i')\phi(i = i'). \quad (11)$$

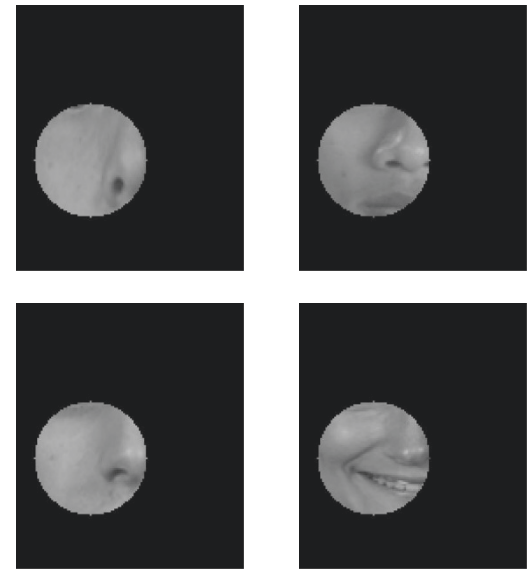


Figure 4: Adapting observation distributions: neighborhood regions. In this example, multiple images from S are given for a given observation distribution adapted to the location of the center patch in the circle.

3.2. Adapting observation function ϕ

The baseline ϕ is modeled here using a nonparametric distribution instead of, say, a Gaussian mixture model (GMM); as indicated above, for each patch i, only the N closest patches i' are chosen from S. While S is a database limited only to face images, there is still variation within faces. That is, a patch's appearance will differ depending upon whether it represents skin, an eye, the mouth, hair, and so forth. So, it is possible that when enhancing a patch $\widehat{M_P^\star}[i]$ from, say, the eye region, that the top N patches selected for ϕ will actually be from, say, the mouth region of the samples in S. This can potentially bring the undesired effect of enhancing the eye in such a way that it resembles the texture of the mouth (see the examples in [7]).

So, even though ϕ already incorporates a strong prior for a whole face image, we propose adapting it on the local level. That is, depending upon a patch's region in the face image, it will be adapted to contain more relevant information:

$$\phi \longrightarrow \phi_i. \quad (12)$$

So, the samples from S used to model ϕ_i can vary from those used to make ϕ_j. In this paper, we propose three ways that ϕ_i can be adapted in a region-dependent way:

(i) neighborhood regions (Figure 4),
(ii) hand-labeled regions (Figure 5),
(iii) learned regions (clusters) (Figure 6).

(a) (b)

Figure 5: Adapting observation distributions: eye/non-eye regions. In this example, eyes, as in (a), have their own observation distribution, built using patch samples from the same regions in S. Likewise, non-eye regions, as in (b), have their own observation distributions, using non-eye regions in the training database S.

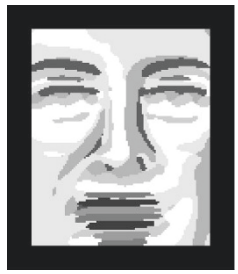


Figure 6: Learned observation clusters with K-means clustering. Shown are the eight regions ($K = 8$), each represented by a different gray-level, of the face learned for this work. The black area represents pixels which do not occur as the center pixel in any patch (cf. Figure 3).

For neighborhood regions, we define a radius distance for each patch i. We then extract patches from S whose center coordinates in their respective, cropped images fall within that distance (in our case, 32 pixels) from the center pixel of i. The motivation for this is that patches near a given patch in the face tend to have the same texture.

Alternately, we can tie distributions for patches together so that a group of patches shares the same distribution:

$$\phi \longrightarrow \phi_{G(i)}, \tag{13}$$

where $G(i)$ is the index for the region/group that the patch i belongs to. One simple example of (13) is to separate the face into two regions, as illustrated in Figure 5:

(1) eye region,
(2) other (non-eye region).

We then extract patches from S whose center pixels' coordinates fall within the same region as the center pixel of a given patch. One of the motivations for doing this approach over the neighborhood approach is the realization that there are discontinuities in areas that have similar texture, particularly with the eyes.

Finally, patches can be clustered together using machine learning techniques. We use K-means clustering [13] to assign each patch to one of K clusters. One of the reasons for using K-means clustering is to make the region definitions data-dependent and, hence, better adapted to the actual face data. The clusters are determined by creating long feature vectors of the high-frequency patches across the N training images, with Q being the number of patches extracted from each image (note that there is a shift of only one pixel between patches during the cluster learning):

$$\begin{bmatrix} \widehat{H_P^1}[1](:) & \widehat{H_P^2}[1](:) & \cdots & \widehat{H_P^N}[1](:) \\ \widehat{H_P^1}[2](:) & \widehat{H_P^2}[2](:) & \cdots & \widehat{H_P^N}[2](:) \\ \vdots & \vdots & \vdots & \vdots \\ \widehat{H_P^1}[Q](:) & \widehat{H_P^2}[Q](:) & \cdots & \widehat{H_P^N}[Q](:) \end{bmatrix}, \tag{14}$$

where each row of (14) is a feature vector input into the K-means and (:) is Matlab notation for the vectorized version of a patch. The result is to find a single clustering from S and to use this single clustering in enhancing any new face image. In the experiments for this work, we set $K = 8$ (Figure 6), and for efficiency reasons, only used a subset of S for computing the K regions.

3.3. Adapting transition function ψ

The baseline ψ models the transition of a patch i only with the patches bordering it (the patches are referred to as neighborhood $\mathrm{N}(i)$ of patch i). A given patch i is then (indirectly) dependent upon any nonneighboring patch given $\mathrm{N}(i)$. However, many of the patches in a face image may be strongly correlated with patches a long distance away. We may therefore want to adapt the definition of $\mathrm{N}(i)$ to include long-distance relationships. One type of long-distance "transition" that we can model is related to the vertical line of face symmetry (see Figure 7). As the face is highly symmetrical, features found on one side of the face will typically be found on the other side of the face. For example, if a person has facial hair on the left side of the face, he will likely also have some on the right side of the face; or someone with freckles on one cheek will also likely have them on the other cheek. For long-distance neighbors, (9) will be modified when computing long-distance transitions:

$$\psi_i^{\dagger}(i = i', j = j'') = \exp\left(\frac{\left[d^{\dagger}\left(\widehat{H_P^{n'}}[i'], \widehat{H_P^{n''}}[j'']\right)\right]^2}{\sigma_{\mathrm{Long}\,Ti}}\right), \tag{15}$$

where $d^{\dagger}(\cdot)$ represents the Euclidean distance between the whole of the first patch and the mirror image of the second patch, with an appropriate normalizing $\sigma_{\mathrm{Long}\,Ti}$, as above.

Alternately, the neighborhood of each patch can be defined using machine learning techniques. For each possible pair of patches (i, j) in the face image, the mutual

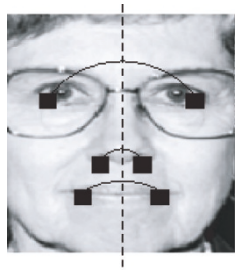


Figure 7: Adapting transition distributions. In this example, three pairs of points are highlighted and connected to illustrate some of the transitions that can be added to a patch's transition distribution, ψ, therefore, is adapted based on its distance from the vertical line of symmetry in a face.

information between the two is

$$MI(i,j) = \sum_{n=1}^{N}\sum_{n'=1}^{N} p\left(\widehat{H_P^n}[i](:), \widehat{H_P^{n'}}[j](:)\right) \cdot \log \frac{p\left(\widehat{H_P^n}[i](:), \widehat{H_P^{n'}}[j](:)\right)}{p\left(\widehat{H_P^n}[i](:)\right) \cdot p\left(\widehat{H_P^{n'}}[j](:)\right)}, \tag{16}$$

but with the simplification that $p(\widehat{H_P^n}[i](:), \widehat{H_P^{n'}}[j](:)) = 0$ when $n \neq n'$, we actually have

$$MI(i,j) \approx \sum_{n=1}^{N} p\left(\widehat{H_P^n}[i](:), \widehat{H_P^n}[j](:)\right) \cdot \log \frac{p\left(\widehat{H_P^n}[i](:), \widehat{H_P^n}[j](:)\right)}{p\left(\widehat{H_P^n}[i](:)\right) \cdot p\left(\widehat{H_P^n}[j](:)\right)}, \tag{17}$$

where the marginal $p(\widehat{H_P^n}[i](:))$ is a single Gaussian $\mathcal{N}(\mu_i, \sigma_i^2)$ with mean and diagonal covariance (denoted using $\mathrm{diag}(\cdot)$), respectively:

$$\mu_i = \sum_{n=1}^{N} \frac{H_P^n[i](:)}{N}, \tag{18}$$

$$\sigma_i^2 = \mathrm{diag}\left(\sum_{n=1}^{N} \frac{(\mu_i - H_P^n[i](:))(\mu_i - H_P^n[i](:))^T}{N-1}\right). \tag{19}$$

$\mathcal{N}(\mu_i, \sigma_i^2)$ is normalized such that

$$\frac{1}{C_i}\sum_{n=1}^{N} p\left(\widehat{H_P^n}[i](:)\right) = 1. \tag{20}$$

The joint $p(\widehat{H_P^n}[i](:), \widehat{H_P^n}[j](:))$ is defined in a similar way. We then identify the learned neighbors of each patch I as those with $MI(i,j) > \delta$, where δ is a global threshold. Figure 8 illustrates some of the learned neighborhoods on a sample training face image. The transition between i and a learned neighbor j is then

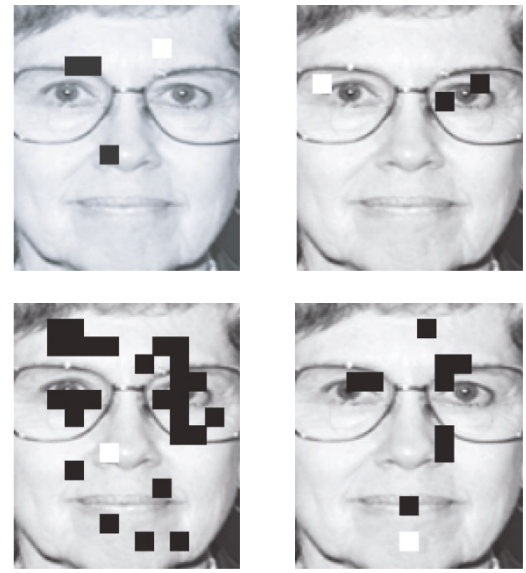


Figure 8: Learning long-distance dependencies. Shown are examples of the long-distance neighbors for a couple selected patches. In each example, the black patches are in the neighborhood of the single white patch. In the work in this paper in learning long-distance dependencies, a patch can have between 0 and 30 learned neighbors.

$$\psi_i^{\dagger\dagger}(i = i', j = j'') = \exp\left(\frac{\left[d^{\dagger\dagger}\left(\widehat{H_P^{n'}}[i'], \widehat{H_P^{n''}}[j'']\right)\right]^2}{\sigma_{\text{Learned } Ti}}\right), \tag{21}$$

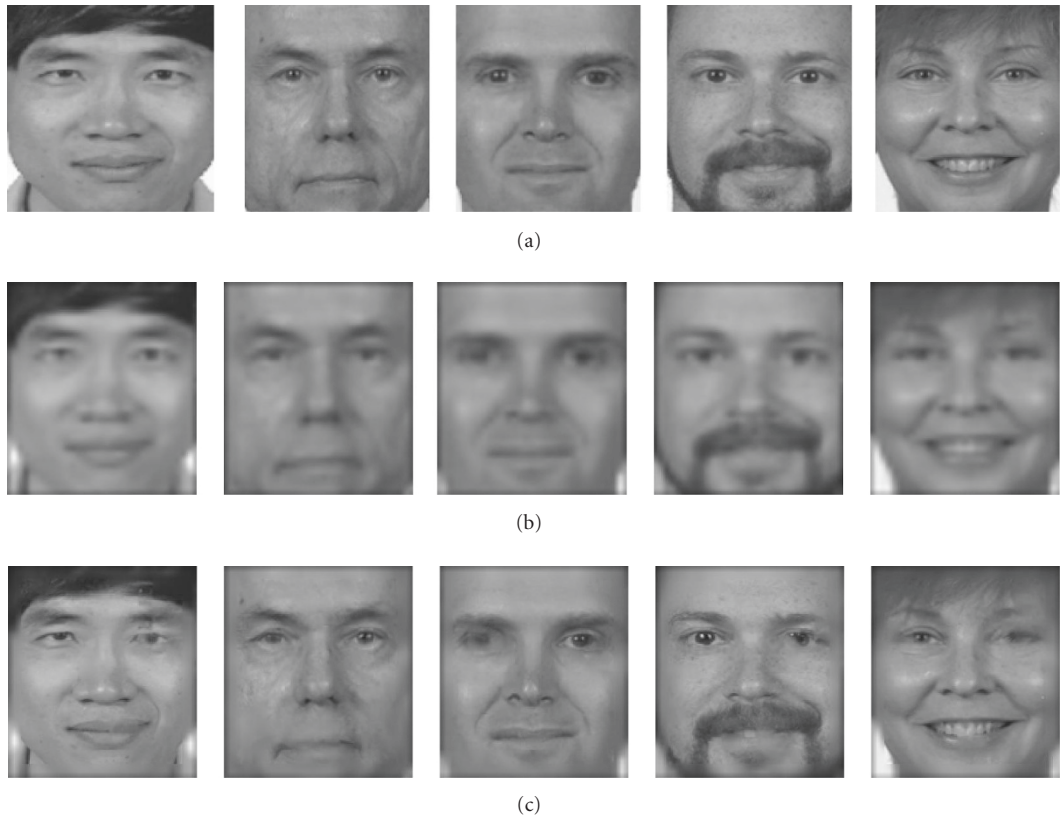
where $d^{\dagger\dagger}$ is the Euclidean distance between the two patches (no mirroring, as done in (15), is performed), with an appropriate normalizing $\sigma_{\text{Learned } Ti}$, like before.

4. FACE ENHANCEMENT EXPERIMENTS

4.1. Setup

In this current work, we are assuming that the face has already been located and properly cropped. We have cropped 1151 faces from the "fa" subset of FERET [14],[1] using the eye and nose coordinates provided with the database. As these are high-quality still images and not low-resolution video images, they are useful for investigating how much of the actual high-frequency we can recover. In future work, we can then investigate their performance in more realistic environments such as surveillance video (though examples on a "real" low-resolution still image are given below in Figures 12, 13, 14, and 15). 951 faces have been randomly extracted for the training set S, while another 100 have been randomly set aside for any tuning of the system

[1] Information on ordering the FERET database can be found at http://www.itl.nist.gov/iad/humanid/feret/.



(a)

(b)

(c)

Figure 9: Baseline results. Row (a) contains the target high-resolution image, while rows (b) and (c) present the bicubic interpolation and baseline MRF results, respectively. Compare with Figures 10 and 11.

Table 2: Super-resolution of 38×33 pixel images to 150×130 pixels (level G_2 to level G_0) showing MSE for the whole image. Results are given using bilinear interpolation, a standard MRF [4], and five of our proposed models: an MRF with observation functions adapted to the region-dependent functions for the eye and non-eye regions; an MRF adapted to the region-dependent functions of neighborhoods; an MRF with observation functions adapted to the region-dependent functions learned using K-means clustering; an MRF with adapted, symmetrical transitions; and an MRF with long-distance, mutual-information-based transitions. As the various MRFs attempt to further enhance low-resolution images that have already been partially enhanced using bilinear interpolation, the percent reduction ("Red.") is with respect to bilinear interpolation. The bicubic interpolation MSE is also given for comparison; the MRFs could potentially do even better in future work if they were enhancing images already partially enhanced using bicubic interpolation. Results are from all 100 images in our test set. As the original, high-resolution images are 150×130 pixels each, the 38×33 pixel images were magnified before enhancement by slightly under a factor of four in each dimension; this was done so as to keep all images used in the algorithm the same size.

Enhancement method	MSE	Red.
Bilinear interpolation	58.9	—
Baseline MRF	54.3	7.7%
MRF: $\phi_{G(i)}$ adapted to eye	52.1	11.5%
MRF: ϕ_i adapted to neighborhood	50.9	13.6%
MRF: $\phi_{G(i)}$ adapted using K-means	53.2	9.7%
MRF: ψ_i adapted to symmetry	53.7	8.8%
MRF: ψ_i adapted using mutual info.	64.3	−9.2%
Bicubic interpolation	49.3	16.3%

and the remaining 100 for testing the system. Each image only appears in one of the lists, but, as many of the 694 subjects appear more than once in the database, a subject can appear on more than one list. Each cropped face is, at high resolution, 150×130 pixels. For experimenting with super-resolution, low-resolution versions of these images have also been produced, as discussed in Section 2.1, by blurring the high-resolution images and subsampling them to produce level G_2 of the Gaussian pyramid, which has images of size 38×33.

In these current experiments, we are only investigating the enhancement of a single image, not of video. In

(d)

(e)

(f)

Figure 10: MRF results: adapting ϕ. Row (10(d)) presents results using $\phi_{G(i)}$ adapted to the eye regions. Row (10(e)) presents results using ϕ_i adapted to neighborhood regions (using a radius around the patch's center pixel). Row (10(f)) presents results using $\phi_{G(i)}$ adapted to regions learned by K-means clustering. Compare with the baseline MRF, which is in row (c) of Figure 9, and with Figure 11. For the subject in column 1, note, for example, (in comparison with the baseline row (c) in Figure 9) the sharper right eye with a clearer boundary in row (10(e)). For the subject in column 2, note that the left eye in row (10(e)) is shinier. For the subject in column 3, note the more realistic eye and better illuminated cheeks in row (10(e)). For the subject in column 4, note the clearer right eye in row (10(e)) and the better illuminated eye in row (10(f)). For the subject in column 5, note in row (10(f)) both the sharper right eye that is consistent with the left eye and the increased detail in the teeth.

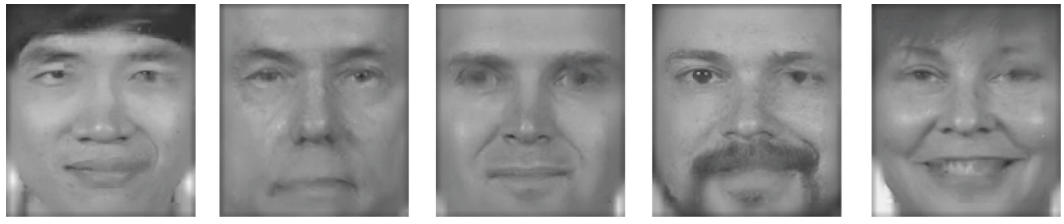
Table 3: Super-resolution of 38×33 pixel images to 150×130 pixels (level G_2 to level G_0) showing MSE for *eye* region (Figure 5(a)). See Table 2 for additional descriptions of the table.

Enhancement method	MSE	Red.
Bilinear interpolation	95.2	—
Baseline MRF	85.1	10.7%
MRF: $\phi_{G(i)}$ adapted to eye	78.6	17.4%
MRF: ϕ_i adapted to neighborhood	77.6	18.5%
MRF: $\phi_{G(i)}$ adapted using K-means	81.6	14.3%
MRF: ψ_i adapted to symmetry	83.7	12.1%
MRF: ψ_i adapted using mutual info.	99.7	−4.7%
Bicubic interpolation	78.9	17.1%

Table 4: Super-resolution of 38×33 pixel images to 150×130 pixels (level G_2 to level G_0) showing MSE for the *non-eye* region (Figure 5(b)). See Table 2 for additional descriptions of the table.

Enhancement method	MSE	Red.
Bilinear interpolation	46.8	—
Baseline MRF	44.1	5.8%
MRF: $\phi_{G(i)}$ adapted to eye	43.3	7.5%
MRF: ϕ_i adapted to neighborhood	42.0	10.3%
MRF: $\phi_{G(i)}$ adapted using K-means	43.8	6.5%
MRF: ψ_i adapted to symmetry	43.8	6.5%
MRF: ψ_i adapted using mutual info.	52.6	−12.3%
Bicubic interpolation	39.5	15.7%

such investigations, we compare our results with those using the approach of [4], which is also concerned with enhancing a single image using MRFs. We do not make direct comparisons to approaches, such as [3] or the main results in [7], that utilize multiple images to produce a single resolved image; this is reserved for future work. Using information only from level 2 of the Gaussian pyramid, we use a baseline MRF from the approach of [4] to infer the high-frequency components missing from the high-resolution G_0 image and indicate how much this baseline MRF compares to using just bilinear (or bicubic, as indicated) interpolation.



(g)

(h)

Figure 11: MRF results: adapting ψ. Row (11(g)) presents results using ψ_i adapted using symmetry in the face. Row (11(h)) presents results using ψ_i adapted using mutual information of the patches. Compare with the baseline MRF, which is in row (c) of Figure 9, and with the MRFs adapting ϕ in Figure 10. In general, the current methods of adapting ψ do not give as much improvement, by themselves, than adapting ϕ.

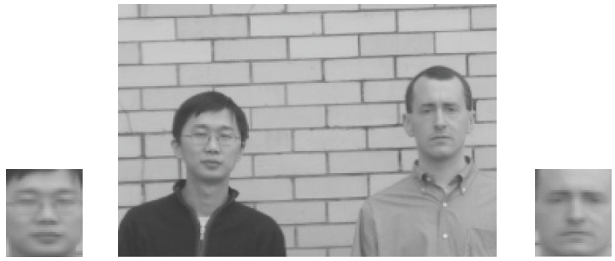
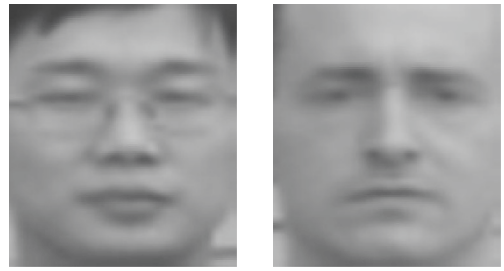
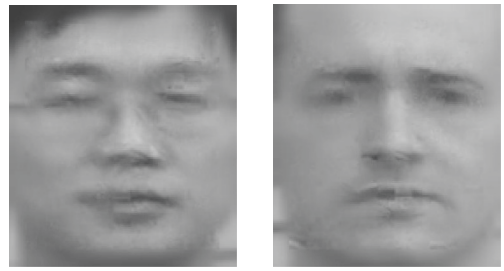


Figure 12: Real low-resolution image, captured by a Canon PowerShot G5, using the 640×480 resolution mode (only 240×160 pixels are cut out from the image and shown here) with the extracted faces (cropped with bilinear interpolation in this figure). The super-resolution results are given in Figures 13, 14, and 15 for the baseline MRF, the MRFs adapted to ϕ_i and to $\phi_{G(i)}$, and the MRFs adapted to ψ_i, respectively.



(b)

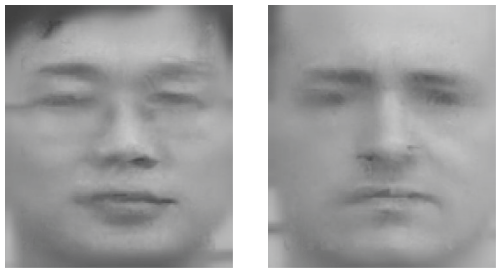


(c)

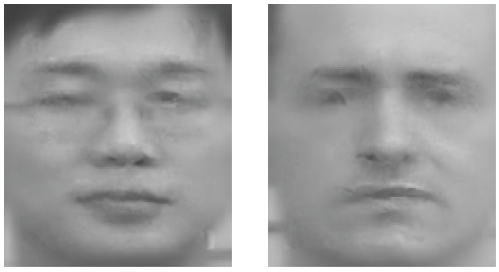
Figure 13: Baseline results on real low-resolution images. Rows (13(b)) and (13(c)) present the bicubic interpolation and baseline MRF results, respectively. To ease comparison with the FERET images of Figure 9, the labeling starts with (13(b)) as no high-resolution images are available.

4.2. Results

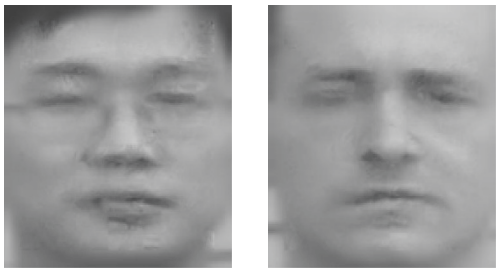
First, to justify the need for having a full-MRF instead of just local observation functions, we show in Table 1 the difference that having transition functions also included between neighboring patches provides. By including ψ with ϕ and having a standard, baseline MRF, we get a mean squared error (MSE) of 54.3. This is an improvement over using either bilinear interpolation or ϕ alone. Given this baseline result using a standard MRF, we then applied our proposed adaptation techniques. Table 2 gives results of the different approaches for enhancing images from G_2, that is, those images which are being enlarged by a factor of approximately 4 in each dimension and then enhanced by super-resolution (note that the MSE values given in this paper do not take into account the unenhanced pixels on the edges of the images, where no high-resolution patches are placed—see Figure 3). Here we see that we can, on average, improve the resolution of the face by using MRFs whose ϕ_i, $\phi_{G(i)}$, or ψ_i function is adapted as indicated (with the exception of adapting ψ_i using mutual information). This is most notable with ϕ_i adapted to



(d)

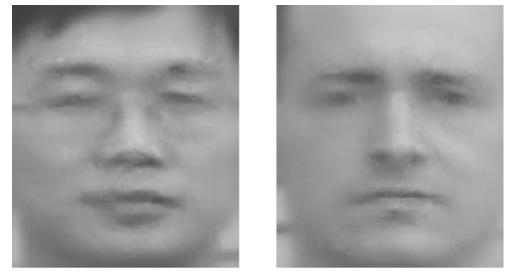


(e)

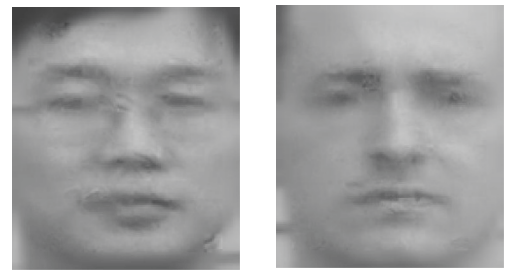


(f)

Figure 14: Adaptive MRF results on real low-resolution images. Row (14(d)) presents results using $\phi_{G(i)}$ adapted to the eye regions. Row (14(e)) presents results using ϕ_i adapted to neighborhood regions (using a radius around the patch's center pixel). Row (14(f)) presents results using $\phi_{G(i)}$ adapted to regions learned by K-means clustering. Compare with the baseline MRF, which is in row (13(c)) of Figure 13, and with Figure 15.



(a)



(b)

Figure 15: Adaptive MRF results on real low-resolution images. Row (15(a)) presents results using ψ_i adapted using symmetry in the face. Row (15(b)) presents results using ψ_i adapted using mutual information of the patches. Compare with the baseline MRF, which is in row (13(c)) of Figure 13 and with Figure 14.

its neighborhood, which reduced the MSE of bilinear interpolation by 13.6%, as opposed to just 7.7% for the baseline MRF. As this method takes patches from S based only upon their distance between their coordinates and the coordinates of the patch being enhanced, this is one of our simpler adaptation techniques. While simple, this technique proves effective in doing example-based super-resolution in a region-dependent manner.

Figures 9, 10, and 11 give the baseline results, results for adapting ϕ_i, and the results for adapting ψ_i, respectively, for some of the images that benefited from the adaptation techniques (any improvements typically came from adapting ϕ_i and $\phi_{G(i)}$ instead of adapting ψ_i). While subjective, the best enhanced image for each of the subjects in Figure 10 is often that of row (10(e)), which are the outputs of adaptive MRFs with ϕ_i adapted to its neighborhood; this is also the adaptive MRF that performed best quantitatively in Table 2. Furthermore, even though they are not tailored specifically to eye/non-eye regions, they also do better when looking specifically at these regions. As the visual improvements are often in the eye regions, we examined the MSE in the images looking only at pixels in the eye regions and also at pixels only in the non-eye regions (as defined by Figure 5, see Tables 3 and 4). Table 3 shows how the modest improvements of Table 2 become even better when looking specifically at the eyes. This could possibly be due to the MRF's concentrating at modeling high-frequency information and to the eyes' containing some of the highest-frequency information in the face (see, e.g., Figure 2(d)). Table 4 shows that the non-eye region of the face, typically with lower-frequency information, benefits less from an adaptive MRF.

In addition to the qualitative results shown in Figures 9, 10, and 11 and the related quantitative results shown in Tables 2, 3, and 4, we also tested our algorithm on real low-resolution images (i.e., those not generated from high-resolution images). Some results are shown in Figures 12, 13, 14, and 15. The quality of these enhanced images could potentially be improved through using a training set S that better matches their domain (e.g., using images with outdoor lighting for S).

5. CONCLUSION

We have proposed a class of adaptive MRFs for increasing the performance of standard example-based super-resolution. By adapting the observation and transition functions to local regions, we restricted the likely high-frequency patches

available for the super-resolution; we showed how doing so not only reduces the MSE associated with a standard MRF but how using such adaptation can produce sharper images.

The next steps in this work of adapting MRFs include improving the modeling of where ϕ_i, $\phi_{G(i)}$, and ψ_i are adapted using machine learning techniques. While using K-means clustering produced acceptable results in adapting $\phi_{G(i)}$, using mutual information in adapting ψ_i can hurt the resolution. The reason for this may lie, in part, in how the adapted ψ_i is defined in (21), which is based on the Euclidean distance between the learned, long-distance neighbors. As the mutual information was based on the joint distribution $p(\widehat{H}_P^n[i](:), \widehat{H}_P^n[j](:))$ (and not on their distance) in (17), it may be more appropriate to use this joint distribution for computing ψ_i.

ACKNOWLEDGMENTS

This research was supported by funding from the IC Postdoctoral Research Fellowship Program. The anonymous reviewers as well as Datong Chen, David Liu, Simon Lucey, Kate Shim, Ted Square, and Wende Zhang were also of assistance in this research.

REFERENCES

[1] A. K. Jain, *Fundamentals of Digital Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1989.

[2] R. Y. Tsai and T. S. Huang, "Multiframe image restoration and registration," in *Advances in Computer Vision and Image Processing*, vol. 1, chapter 7, pp. 317–339, JAI Press, Greenwich, Conn, USA, 1984.

[3] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Transactions on Image Processing*, vol. 5, no. 6, pp. 996–1011, 1996.

[4] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 25–47, 2000.

[5] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, vol. 19 of *Computer Science Workbench Series*, Springer, Tokyo, Japan, 2001.

[6] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Fransisco, Calif, USA, 1988.

[7] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.

[8] G. Dedeoğlu, T. Kanade, and J. August, "High-zoom video hallucination by exploiting spatio-temporal regularities," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. 151–158, Washington, DC, USA, June–July 2004.

[9] S. L. Lauritzen, *Graphical Models*, Oxford Statistical Science Series, No. 17, Clarendon Press, Oxford, UK, 1996.

[10] H. W. Schneiderman, *A statistical approach to 3D object detection applied to faces and cars*, Ph.D. thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, Pa, USA, May 2000, CMU-RI-TR-00-06.

[11] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[12] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Generalized belief propagation," in *Proceedings of Advances in Neural Information Processing Systems 13 (NIPS '00)*, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds., vol. 13, pp. 689–695, MIT Press, Cambridge, Mass, USA, 2001.

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

[14] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

Todd A. Stephenson has been a Senior Scientist at ReallaeR, LLC, of Saint Leonard, Maryland, since November 2005. From April 2004 to November 2005 he was a Postdoctoral Fellow at the Advanced Multimedia Processing Laboratory, Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania. From March 1999 to June 2003 he was a Research Assistant in the speech processing group of the IDIAP Research Institute in Martigny, Switzerland. From June 1995 to August 1997 he worked in the Consumer Markets Division of AT&T Corporation in Piscataway, New Jersey, and in Somerset, New Jersey. Todd received the Ph.D. degree in electrical engineering from the Swiss Federal Institute of Technology Lausanne (EPFL) in 2003, the M.S. degree in cognitive science and natural language from the University of Edinburgh in 1998, and the B.S. degree in Mathematics from the Pennsylvania State University in 1995. His research interests include computer vision, machine learning, speech recognition, and natural language processing. He is a Member of the IEEE.

Tsuhan Chen has been with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania, since October 1997, where he is currently a Professor. He directs the Advanced Multimedia Processing Laboratory, working on multimedia signal processing, biometrics, computer vision, and computer graphics. From August 1993 to October 1997, he worked at AT&T Bell Laboratories, Holmdel, New Jersey. Tsuhan helped create the Technical Committee on Multimedia Signal Processing and the Multimedia Signal Processing Workshop, in the IEEE Signal Processing Society. He was appointed the Editor-in-Chief for IEEE Transactions on Multimedia for 2002–2004. He also served as Associate Editor for IEEE Transactions on Circuits and Systems for Video Technology, IEEE Transactions on Image Processing, IEEE Transactions on Signal Processing, and IEEE Transactions on Multimedia. He coedited a book titled *Advances in Multimedia: Systems, Standards, and Networks.* Tsuhan received the B.S. degree in electrical engineering from the National Taiwan University in 1987, and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, California, in 1990 and 1993. He received the Charles Wilts Prize for outstanding independent research. He was a recipient of the National Science Foundation CAREER Award from 2000 to 2003.