Silhouette estimation

Richard G. Paxman,[1,*] David A. Carrara,[1] Paul D. Walker,[1] and Nicolas Davidenko[2]

[1]*General Dynamics Advanced Information Systems, P.O. Box 990, Ypsilanti, Michigan 48197, USA*
[2]*Psychology Department, UC Santa Cruz, 1156 High Street, Santa Cruz, California 95064, USA*
**Corresponding author: rick.paxman@gd-ais.com*

Received February 13, 2014; revised April 8, 2014; accepted April 9, 2014;
posted May 23, 2014 (Doc. ID 206333); published June 30, 2014

Silhouettes arise in a variety of imaging scenarios. Pristine silhouettes are often degraded via blurring, detector sampling, and detector noise. We present a maximum *a posteriori* estimator for the restoration of parameterized facial silhouettes. Extreme dealiasing and dramatic superresolution, well beyond the diffraction limit, are demonstrated through the use of strong prior knowledge. © 2014 Optical Society of America

OCIS codes: (100.3010) Image reconstruction techniques; (100.3008) Image recognition, algorithms and filters; (100.1830) Deconvolution.
http://dx.doi.org/10.1364/JOSAA.31.001636

1. INTRODUCTION

A. Silhouettes in Imaging

A silhouette is the image of an object represented as a solid shape of a single color (typically black) so that its edges match the object's outline. The term derives from the Frenchman Étienne de Silhouette, whose short tenure in 1759 as finance chief caused him to become the object of public ridicule. His austerity policies led to the expression *à la Silhouette*, in reference to things perceived as cheap. The term was applied to the contemporaneous art form of cutting out black paper to create a facial profile, thus serving as an inexpensive alternative to a painting [Fig. 1(a)]. Although the term silhouette often refers to facial profiles, it is also commonly applied to the binary representation for the shape of any object.

Silhouettes arise in a variety of imaging scenarios. These include images of shadows that are cast on either a uniform or a nonuniform but known background, as illustrated in Fig. 1(b). Silhouettes also occur when an object occludes a known background. This case is particularly evident when a bright background, such as the sun or the moon, is occluded by a relatively dark object, such as a satellite or an aircraft, as illustrated in Fig. 2.

Silhouette images are often degraded through multiple mechanisms, including blur, detector sampling (e.g., box-car sampling characteristic of standard digital cameras), and noise. We desire to restore a crisp silhouette from a degraded image for use in estimation, classification, and identification tasks. For example, the restored silhouette of a satellite can be used for pose estimation. Davidenko has quantified the classification performance (e.g., into classes of gender, age, and race) of observers looking at standardized facial silhouettes [1]. Restored silhouettes of a common object collected over a range of orientations can be used to reconstruct the object's three-dimensional (3D) visual hull [2,3], also useful in identification and classification tasks. Clearly, facial silhouettes can be used for identifying an individual from a gallery of candidates. In addition, the dynamic silhouette of a walker can be used in gait recognition [4]. Because we are primarily interested in the behavior of the silhouette in this case, we can consider image-compression and compressive-sensing methods that ignore the detail within the silhouette (Fig. 3).

When silhouettes can be well estimated, they can subsequently be used to constrain adjacent estimation problems. Silhouettes can fill the role of a support (or indicator) function that can be used for phase retrieval [5], or for superresolution [6].

B. Scope of Paper

In this paper we concentrate on estimating facial silhouettes because of the availability of a suitable statistical database. However, the estimators derived herein can be extended to a wide variety of silhouette-estimation problems. An efficient spline parameterization for facial silhouettes, along with an associated statistical model for that parameterization, is presented in Section 2. In Section 3, we present a continuous-discrete (CD) data-acquisition model in which a silhouette, defined on the continuous real plane, is blurred, discretely sampled, and corrupted with noise. Section 4 presents candidate silhouette estimators. A maximum-likelihood (ML) estimator is shown to work well for data corrupted with modest amounts of noise. However, a maximum *a posteriori* (MAP) estimator, first presented in [7], provides needed management of noise amplification when noise levels are more substantial. We also present a practical MAP estimator that accommodates misregistration between the data and the model as well as acquisition over finite fields of view. Estimator performance is discussed in Section 5, where extreme dealiasing and dramatic superresolution (restoration of spatial frequencies beyond the cutoff frequency) are demonstrated. These dramatic results are attributed to the use of strong prior knowledge: that the object is binary, that the silhouette can be characterized by only a small number of parameters, and that the statistical distribution of the parameters is known. Restoration of degraded silhouettes with such prior knowledge is closely related to use of an opacity constraint for phase retrieval [8–10] and for superresolution [11]. We conclude

Fig. 1. Example silhouettes. (a) Classic facial silhouette constructed as a paper cutout. (b) Silhouette created by casting a shadow onto an approximately uniform background. Note the information in the shadow that is not evident in the direct image of the object.

in Section 6 with recommendations for future research directions.

2. PARAMETERIZATION AND STATISTICAL MODEL FOR FACIAL SILHOUETTES

Two-dimensional (2D) images are most often parameterized with pixel values. Silhouettes can be parameterized with pixels, but since the image is a binary-valued connected shape, pixel parameterization is inefficient. A silhouette is defined by its boundary, which is a one-dimensional curve embedded in a 2D space, sometimes referred to as having 1.5 dimensions. Therefore parameterizing a silhouette image with a spline parameterization can be far more efficient than using a pixel parameterization. Davidenko developed a spline parameterization that can be used for facial silhouettes in profile [1]. We briefly review Davidenko's method. Figure 4 shows the sequence for creating a spline-parameter representation from a gray-level facial profile. First, 18 landmark points, defined in words, are manually located. Examples of landmark-point descriptions include the tip of the nose (maximum-curvature point), the bridge of the nose, and half-way (in arc length) between the tip and the bridge of the nose. Once these landmark points are located, then the array of points is rotated and scaled so that the first and last landmarks fall at coordinates (0,0) and (0,1), respectively. These normalized landmark-point arrays are then interpolated with a cubic spline that forces a cusp at the intersection of the lips (one of the landmark points). A vertically elongated rectangle is then added, and the interior of the silhouette is set to black. For comparison, the figure also shows an image of the original gray-scale representation that has been converted to a binary representation with a threshold operation. This comparison suggests that Davidenko's simple spline-fit strategy provides a reasonable representation of the facial silhouette.

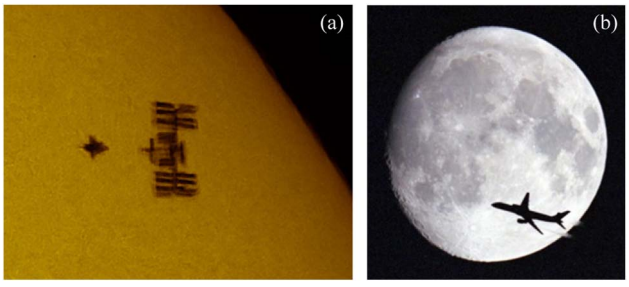

Fig. 2. Silhouettes formed by relatively dark objects occluding a bright background. (a) Solar transit of the Space Shuttle and the International Space Station, image courtesy of Thierry Legault, http://www.astrophoto.fr. (b) Aircraft occluding the moon, image courtesy of Kurt Gleichman.

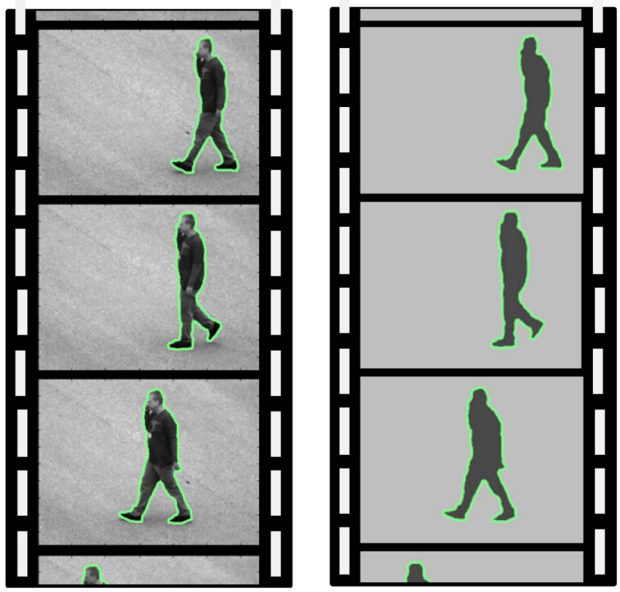

Fig. 3. Multiple frames of silhouettes of a walker occluding a background. Such silhouettes can be used for gait recognition. Silhouettes enable image compression or compressive sensing.

The procedure illustrated in Fig. 4 was carried out for 384 facial-profile images derived from the FERET database [12,13]. This resulted in a 32-element (x and y values for each of 16 variable landmark points, after normalization) parameter vector, α, for each of the 384 faces in the database. These parameter vectors can be viewed as highly correlated random vectors. Each element in the parameter vector was found to be approximately normally distributed [1]. Accordingly, we model the statistics of the 32-element parameter vector, α, with a multivariate normal probability distribution as follows:

$$p_\alpha(\alpha) = [(2\pi)^M \det(\hat{K}_\alpha)]^{-1/2} \exp\left[-\frac{1}{2}(\alpha-\bar{\alpha})^{\mathrm{T}}\hat{K}_\alpha^{-1}(\alpha-\bar{\alpha})\right] \quad (1)$$

$$= [(2\pi)^M \det(\hat{K}_\alpha)]^{-1/2} \exp\left[-\frac{1}{2}[\mathrm{A}^{\mathrm{T}}(\alpha-\bar{\alpha})]^{\mathrm{T}}\Lambda^{-1}[\mathrm{A}^{\mathrm{T}}(\alpha-\bar{\alpha})]\right], \quad (2)$$

where M is the dimension of the parameter vector, $\hat{K}_\alpha$ is the parameter-vector sample covariance, $\bar{\alpha}$ is the parameter-vector sample mean, the superscript T indicates transpose, A is a matrix constructed with principal-component vectors (eigenvectors of the sample covariance) in its columns, and Λ is a diagonal matrix with corresponding principal-component values (eigenvalues of the sample covariance) for the diagonal elements. Under this model, a facial silhouette is characterized with a mere 32 parameters (far more efficient than with a pixel parameterization), and the statistics of the parameters are given by a multivariate normal distribution.

It is instructive to plot the principal-component spectrum in order of descending value, as shown in Fig. 5. The principal-component values have been normalized to indicate the portion of the total variance associated with the corresponding principal-component vector. For example, the first principal component accounts for roughly 42% of the variance in the

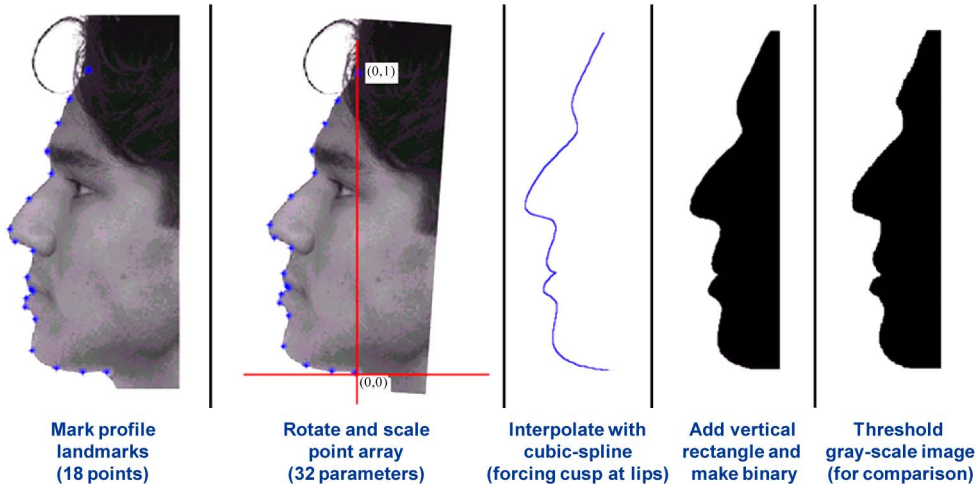

Fig. 4. Steps in parameterizing a facial silhouette from a gray-level profile image.

384 facial silhouettes. The rapid decay of the spectrum suggests that the number of parameters needed to characterize a facial silhouette may be considerably less than 32, depending upon the task.

3. DATA MODEL

We are interested in estimating a crisp silhouette from data for which the silhouette has been blurred, sampled using a camera detector, and corrupted with noise. The CD data-acquisition model [14] is given by

$$d(\mathbf{x}) = D_{\mathbf{x}}\left\{\int f(x',\alpha)s(x-x')\mathrm{d}x'\right\} + n(\mathbf{x}) \tag{3}$$

$$= g(\mathbf{x}) + n(\mathbf{x}), \tag{4}$$

where f is the 2D binary silhouette, parameterized by the vector α and defined on the continuous real plane, s is the known blur function, and $D_{\mathbf{x}}\{\cdot\}$ is a discretizing operator (modeling the action of a noiseless detector array) that takes a continuous signal and returns a noiseless vector g defined on the multi-index $\mathbf{x}$. The acquired data, d, has been corrupted by additive detector noise, n. When a shadow is cast directly onto a detector array by an extended illuminator, the blur function corresponds to the illuminator distribution function. Alternatively, a camera may be used to record an image of a shadow cast by a point illuminator onto a uniform surface. In this case the blur is due to the point-spread function (PSF) of the camera, arising from camera aberrations and diffraction effects. In most practical applications, the blur will be due to both the illumination extent and the camera PSF. Figure 6 shows the fundamental degradation mechanisms in simulation for the average facial silhouette, $f(x,\bar{\alpha})$, for known blur arising from the extent of the sun and the shadow being cast 1.5 m from the face.

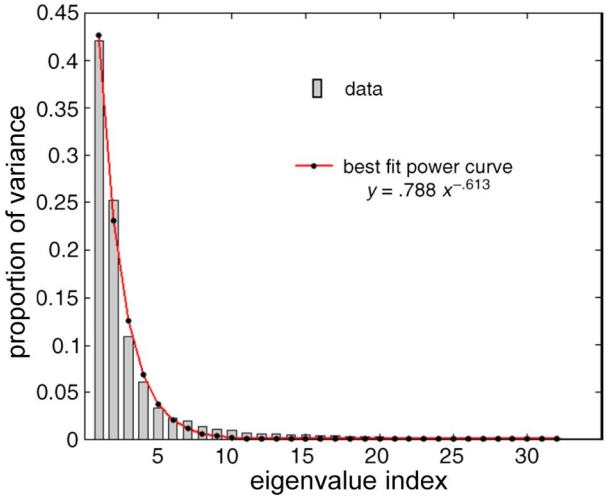

Fig. 5. Principal-component spectrum for parameterized facial silhouettes, normalized to indicate proportion of variance. The red line indicates a power-law fit to the spectrum.

The pristine (crisp) shadow can be defined continuously on a 2D real window (the Cartesian product of two real intervals) through the use of splines, but is depicted on a 512×1024 representation grid that serves as a surrogate for the continuous object. The efficiency of the 32-spline-parameter representation of the facial silhouette can be appreciated by comparison to the more than ½ million (512×1024) pixel parameters needed to represent the silhouette on this representation grid. The continuous blurred image is approximated by discrete convolution with a circular disk that is the geometric pinhole image of the sun that would appear on a screen 1.5 m behind the pinhole. For this geometry, the PSF diameter is 59 representation pixels in diameter. Note that the shadow region will be subject to indirect illumination. We performed a MODTRAN [15] analysis using a representative atmosphere (mid-latitude summer, rural, visibility = 23 km), to determine that the irradiance in the direct-illumination region will be approximately 6.2 times that of the indirect-illumination (shadow) region, found by integrating the sky-shine contribution over a hemisphere. Of course this ratio will vary with geometry, time of day, atmospheric conditions, etc. Noiseless detector sampling occurs when the continuous signal within each of the 32×64 detector elements is spatially integrated. This was approximated by aggregating the 16×16 representation-grid values of the blurred image that fall within each detector element. Finally, independent and identically

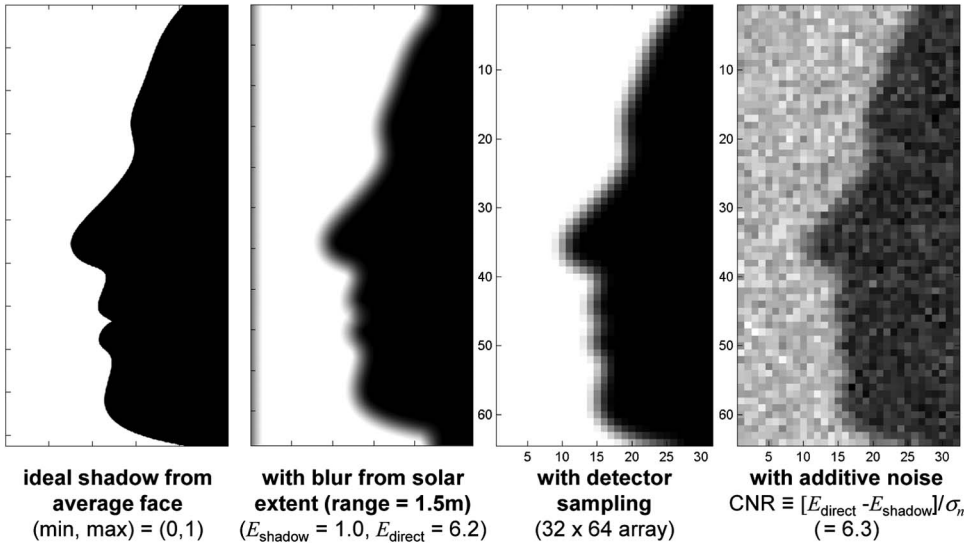

Fig. 6. Mechanisms of degradation in silhouette imaging. From left to right: the ideal continuous silhouette of the average face represented on a 512×1024 grid, blurred by solar extent for shadow 1.5 m away from face, 32×64 discrete array resulting from detector sampling, and with additive read noise.

distributed Gaussian read noise values are added to the sampled image. We characterize the degree of noise corruption with a contrast-to-noise ratio (CNR) metric,

$$\mathrm{CNR} \equiv \frac{E_{\mathrm{direct}} - E_{\mathrm{shadow}}}{\sigma_n}, \tag{5}$$

where E represents irradiance and σ_n represents the standard deviation of the additive noise.

4. ESTIMATORS

In this section, we examine various estimators for use in silhouette estimation.

A. Maximum-Likelihood Estimation

It is well known that for the case of additive Gaussian noise, the ML estimate minimizes the squared error in the predicted data,

$$\hat{\alpha}_{\mathrm{ML}} = \underset{\alpha}{\operatorname{argmin}} \sum_{\mathbf{x}} \left(d(\mathbf{x}) - D_{\mathbf{x}} \left\{ \int f(x';\alpha) s(x - x') \mathrm{d}x' \right\} \right)^2. \tag{6}$$

We use a limited-memory BFGS optimizer [16], starting from a naïve initial estimate (the mean facial silhouette parameter vector), to find an implicit estimate. Bear in mind that the ML estimate is a parameter vector that implies a facial silhouette defined continuously on a 2D real window through the use of splines. The ML-estimation strategy works well for cases involving modest amounts of noise. Figure 7 shows the estimated facial silhouette for the case of CNR = 10. In this experiment, we used a truth silhouette that was a random draw from our facial-silhouette probability distribution function (PDF) and that was clearly distinct from the average facial silhouette. In spite of the corrupting effects of blur, sampling, and noise, the ML estimate is excellent and captures the character of the true silhouette. The quality of the estimate is surprisingly good, given that no regularization strategy has been employed. In the figure we include a difference image that indicates where the estimated silhouette differs from the true silhouette. At most locations there is no difference (zero represented by a mid-gray value). The only differences are within a thin ribbon proximate to the true-silhouette facial boundary. Estimation performance can be quantified by one of many candidate dissimilarity measures that are used to compare two curves (e.g., the Hausdorff distance or the Fréchet distance) [17]. Ultimately, only metrics that can be related to task performance are meaningful. Herein, we use the area of symmetric difference (ASD) between the true and estimated silhouettes, defined over a common region of interest (ROI), and normalize it by the arc length of the true-silhouette facial boundary. This measure has the favorable feature that it satisfies the properties of a metric. In addition, the normalized ASD (NASD) has a relatively simple geometric interpretation: it is the area of all nonzero regions (typically a thin ribbon) in the ROI difference image, normalized by the arc length of the true-silhouette boundary over that same ROI. The units of the NASD are in length, and it can be interpreted as a kind of average displacement from the true-silhouette boundary.

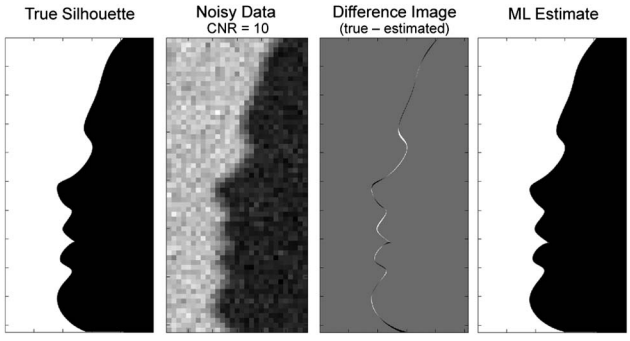

Fig. 7. Maximum-likelihood (ML) estimation performance for CNR = 10. The difference image indicates that the estimate is excellent and the NASD is 1.42 representation-grid pixels. For reference, a single representation pixel is approximately 0.25 mm for an average face size.

In Fig. 8 we provide similar experimental results, except that the amount of noise has been increased and the quality of the data is far more challenging. The ML estimates in these noisy regimes develop caricature-like silhouettes. The ML estimate associated with the particular noise realization at CNR = 3.16 features a "Bob Hope" nose, whereas the estimate at CNR = 1.73 developed "Andy Rooney" eyebrows. These caricature features arise when principal components with small principal-component values (therefore statistically unlikely) are nevertheless useful in minimizing the squared error in the data. The additional noise has rendered both restorations unrecognizable. It is not surprising that we need a strategy to avoid overfitting the data in these noisy regimes.

B. Maximum *a posteriori* Estimation

The MAP estimate for the parameter vector α maximizes the log of the posterior probability density function,

$$\hat{\alpha}_{\mathrm{MAP}} = \underset{\alpha}{\operatorname{argmax}}\{\ln[\mathrm{pr}(d|\alpha)] + \ln[\mathrm{pr}(\alpha)]\} \tag{7}$$

$$\approx \underset{\alpha}{\operatorname{argmin}}\left\{\sum_{\mathbf{x}}\left[d(\mathbf{x}) - D_{\mathbf{x}}\left\{\int f(x';\alpha)s(x-x')\mathrm{d}x'\right\}\right]^2 + \eta[\mathrm{A}^{\mathrm{T}}(\alpha-\bar{\alpha})]^{\mathrm{T}}\Lambda^{-1}[\mathrm{A}^{\mathrm{T}}(\alpha-\bar{\alpha})]\right\}, \tag{8}$$

where we have introduced the hyperparameter η to help control noise amplification. The right-hand side for each of the above equations consists of two terms. The first is a data-agreement term, identical to that used in ML estimation. The second is a regularization term that penalizes components that are statistically unlikely. Figure 9 provides a comparison between ML and MAP estimates for three CNR levels. Clearly, the regularization provided by MAP estimation eliminates the caricature-like features evident in the low-CNR ML estimates. This is done by biasing the estimate toward the average face.

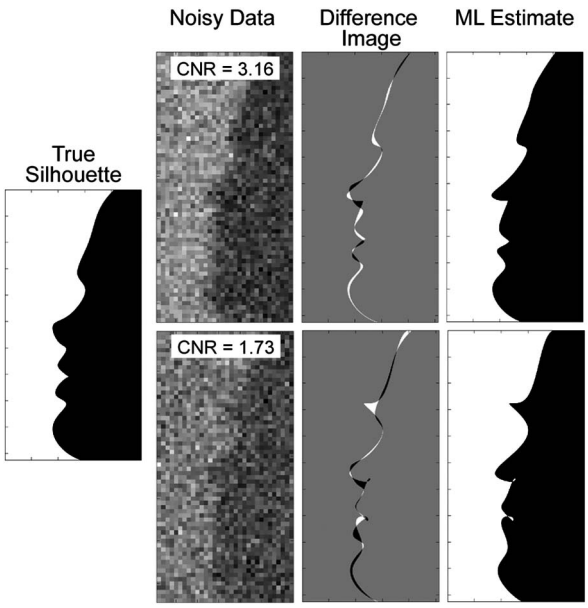

Fig. 8. ML estimation of silhouette for CNR = 3.16 and 1.73. In these low-CNR regimes, the ML estimates develop caricature features that do not match the true silhouette.

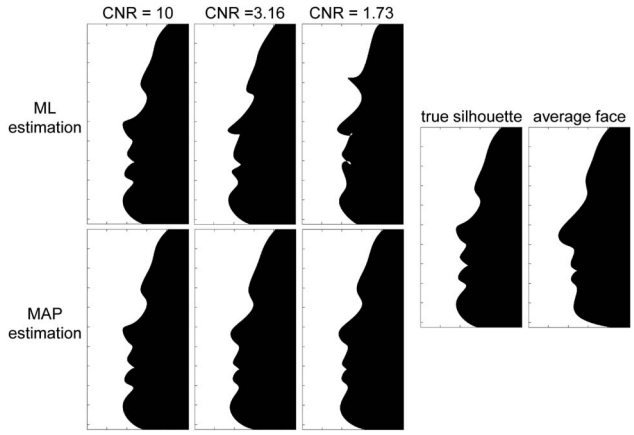

Fig. 9. Maximum-likelihood (ML) and maximum *a posteriori* (MAP) estimates of facial silhouettes for three CNR levels. The data associated with these CNR levels are shown in Figs. 7 and 8. The true silhouette and the average facial silhouette are also provided for comparison.

We have included the average facial silhouette in the figure to illustrate that the MAP estimates more closely resemble the true silhouette than the average face. For these simulations, we selected the hyperparameter value so as to roughly minimize the NASD. In practice, the hyperparameter value can be found using generalized cross validation [18], or related strategies. It is remarkable that the MAP estimate for CNR = 1.73 resembles the true silhouette as well as it does, given the poor quality of the corresponding data shown in Fig. 8. This can be attributed to the use of strong prior knowledge: that the object is binary, that the silhouette can be characterized by only 32 parameters, and that the statistical distribution of the parameters is known.

C. Practical MAP Estimation

All of the estimates reported thus far have assumed that the registration between collected data and model landmark endpoints, (0,0) and (0,1), is known *a priori* and that the field of view (FOV) covers the entire silhouette. In most practical applications, neither of these assumptions holds. The MAP estimator is readily adapted to accommodate these considerations. First, we generalize the silhouette parameter vector to include deterministic in addition to probabilistic parameters:

$$\alpha \equiv [\alpha_p^T, \alpha_d^T]^T, \tag{9}$$

where α_p is the vector of probabilistic parameters (heretofore referred to as α in the absence of deterministic parameters) and α_d is a vector of four deterministic parameters, accommodating lateral registration, rotation, and magnification of the data relative to the normalized-landmark model. A convenient parameterization for the deterministic vector is simply x and y coordinate values of the first and last landmark points.

We also elaborate on our interpretation of the discretizing operator, $D_{\mathbf{x}}\{\cdot\}$, which models the action of a detector array, converting a continuously defined function into a vector, $g(\mathbf{x})$. In addition to modeling the detector-element spatial response and sampling rate, this operator models the FOV governed by the extent of the detector array. Therefore, variable FOV is implicit in the operator notation. With these modifications, the practical MAP estimate is readily found to be

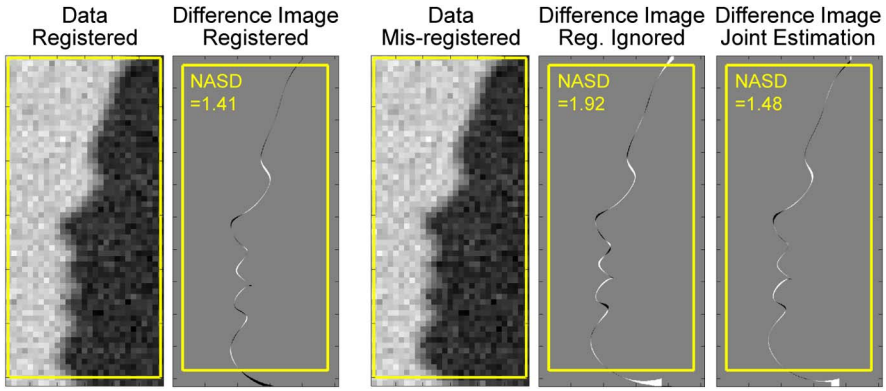

Fig. 10. Demonstration of practical MAP estimation for modest misregistration in the case of using the full FOV. The central panel shows CNR = 10 misregistered data such that the first (below-chin) and last (hairline) landmark points are translated by (−2.0, 1.5) and (0.3, −0.7) detected pixels, respectively. The joint estimate is significantly better than the naïve estimate (misregistration ignored) and comparable to the estimate made with registered data. The yellow borders represent the field of view (FOV) and the ROI over which the NASD (in representation pixels) was computed for the data and the difference images, respectively.

$$\hat{\alpha}_{\mathrm{PMAP}} = \underset{\alpha}{\operatorname{argmin}}\left\{\sum_{\mathbf{x}}\left[d(\mathbf{x}) - D_{\mathbf{x}}\left\{\int f(x';\alpha)s(x-x')\mathrm{d}x'\right\}\right]^2 + \eta[\mathrm{A}^{\mathrm{T}}(\alpha_p - \bar{\alpha}_p)]^{\mathrm{T}}\Lambda^{-1}[\mathrm{A}^{\mathrm{T}}(\alpha_p - \bar{\alpha}_p)]\right\}, \quad (10)$$

where the FOV extent is implicit in the discretizing operator.

We performed a simulation experiment to investigate the ability of practical MAP estimation to accommodate modest misregistration while essentially collecting data over the full facial silhouette. In this experiment, misregistered data were created by transforming the true landmark points (laterally, rotationally, and in magnification) so that the first (below-chin) landmark point was translated by (−2.0, 1.5) detected pixels and the last (hairline) landmark point was translated by (0.3, −0.7) detected pixels. Joint estimation of deterministic as well as probabilistic parameters is shown to perform better than naïve estimation, in which the data and model are assumed to be properly registered and deterministic parameters are not estimated. The results are illustrated in Fig. 10.

We also performed an experiment to investigate the effects of varying the FOV, in isolation from registration effects. The results, shown in Fig. 11, indicate that estimation performance degrades very slowly with decreasing FOV until the FOV is smaller than the ROI, where estimation accuracy degrades rapidly with decreasing FOV. In addition, estimation for regions outside of the FOV is relatively poor, as expected.

Finally, we provide a result involving both misregistration and reduced FOV. In this experiment, the misregistration was much more dramatic and the FOV was relatively small. Whereas the misregistration parameters were not accurately estimated with the practical MAP estimator, the fit to the true silhouette is remarkably good over the FOV, as illustrated in Fig. 12. This suggests that the statistical correlations between

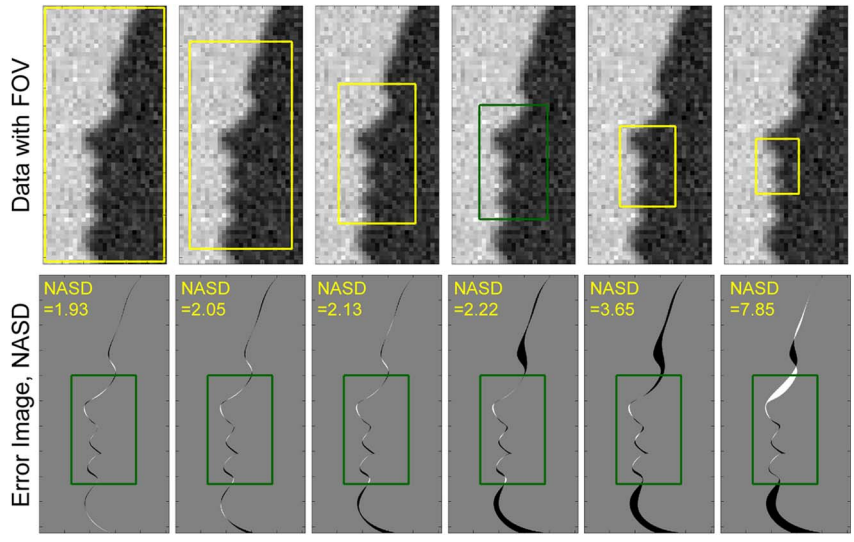

Fig. 11. Estimation as the FOV varies in the case of no registration errors. The upper row shows the data (CNR = 10) and the varying FOVs. The lower row shows the corresponding error images and the NASD values (in representation pixels), computed over a common ROI, equivalent to the fourth largest FOV. The estimates degrade very slowly with decreasing FOV until the FOV is smaller than the ROI, where estimation accuracy degrades rapidly with decreasing FOV.

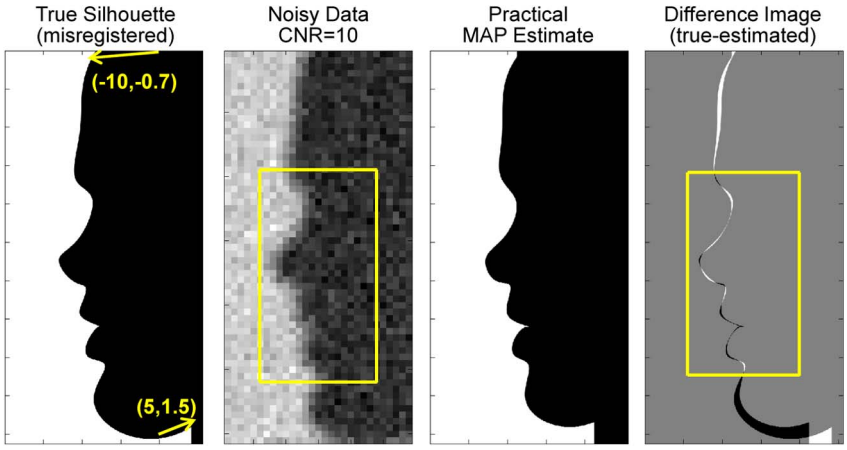

Fig. 12. Practical MAP estimation with dramatic misregistration and reduced FOV. The true silhouette and corresponding data have been misregistered (translated, rotated, and scaled) relative to the model by translating the first and last landmark points, as shown. The rectangular box indicates the FOV over which the data were used for estimation. The estimated facial silhouette is relatively accurate within the region of the FOV, with an NASD of 2.14 representation pixels.

landmark points tend to be local. The results of Fig. 12 give us confidence that MAP estimation of silhouettes can be accomplished in practical settings.

5. ESTIMATION PERFORMANCE

A. Dealiasing

Consider the case of a shadow being cast by an extended illuminator (like the sun) onto a planar detector array such as a common CCD. Under a geometric-optics model, the PSF will be a scaled version of the illuminator distribution. We model the PSF associated with solar illumination as a circular disk or top hat. The modulation transfer function (MTF) for such a PSF is a rectified sombrero (jinc) function. Although this MTF has zeros, the area over which the MTF is zero is of measure zero, and there is no cutoff frequency for this image. Therefore any sampled data will technically be aliased. In our simulation to demonstrate dealiasing, we approximated a crisp facial silhouette with a 256×512 representation grid, as illustrated in Fig. 13. The discrete Fourier transform (DFT) of this image shows that there is spatial-frequency content distributed out to the edge of the Fourier array. This is true of the blurred image as well, although the content has been modulated down owing to the associated MTF. Blur from a circular disk imposes a sombrero-function modulation on the Fourier representation. Box-car sampling can be modeled with convolution

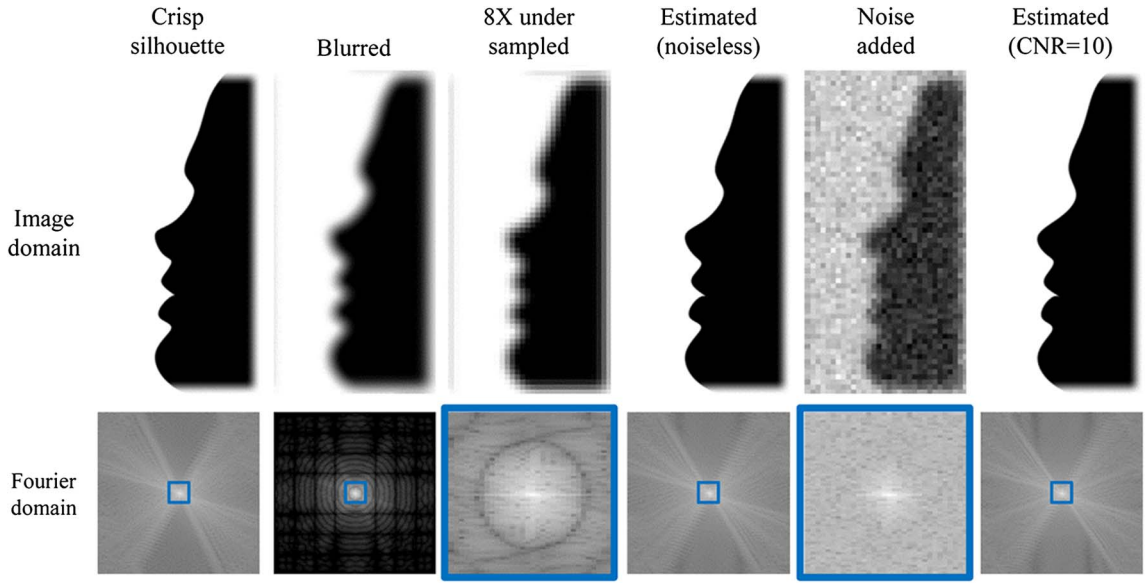

Fig. 13. Demonstration of extreme dealiasing enabled by strong prior knowledge. The crisp silhouette is constructed on a 256×512 representation grid, and the corresponding Fourier-domain representation (magnitude of the DFT, displayed with log transformation and scaling, common for all panels, selected to emphasize relevant features) shows that object spatial frequencies are nonzero out to the edge of the array. Blur from a circular disk imposes a sombrero-function modulation on the Fourier representation. Additional blur from the 8×8 aggregation kernel imposes an additional 2D sinc modulation. The bold squares (in blue) represent the Nyquist frequency for the undersampled data, equal to 1/8th the Nyquist frequency for the crisp silhouette. The Fourier representation of the undersampled data (enlarged in the figure by a factor of 8) is a mixture of many overlapping Fourier components. The silhouette estimated from the noiseless undersampled data is virtually indistinguishable from the true (crisp) silhouette. When noise is added, silhouette estimation degrades gracefully.

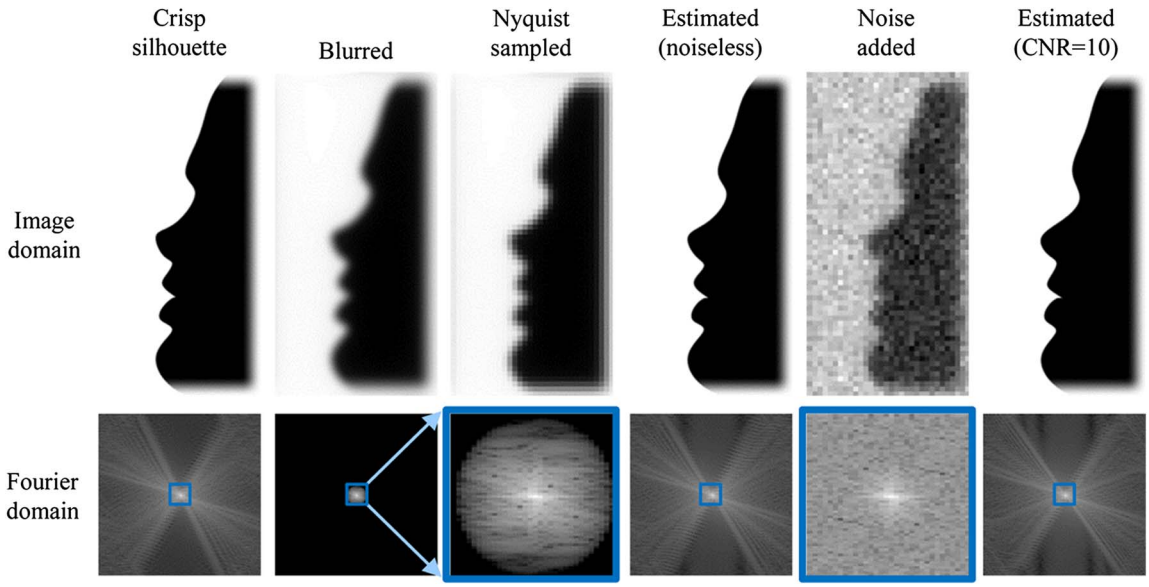

Fig. 14. Demonstration of dramatic superresolution enabled by strong prior knowledge. Blur is caused by imaging with an unaberrated optical system and convolving with an 8×8 aggregation filter. Accordingly, the Fourier representation of the sampled data (enlarged in the figure by a factor of 8) displays a clear cutoff frequency. The bold squares (in blue) represent the Nyquist frequency for the sampled data, equal to 1/8th the Nyquist frequency for the crisp silhouette. The silhouette estimated from the noiseless sampled data is virtually indistinguishable from the true (crisp) silhouette. When noise is added, silhouette estimation degrades gracefully.

with an aggregation kernel followed by decimation. Accordingly, additional blur from an 8×8 aggregation kernel imposes an additional 2D-sinc modulation on the blurred image. Decimation of the blurred image to 32×64 detected pixel values corresponds to undersampling by a factor of 8 and mixing of many (64) overlapping Fourier components so that the Fourier data are severely aliased.

Figure 13 illustrates that the estimated silhouette successfully disentangles the extreme aliasing when a single frame of data is undersampled by a factor of 8. This is accomplished through the use of strong prior knowledge. Note that classical multiframe dealiasing would require 64 frames, with carefully orchestrated subpixel shifts, to perform the dealiasing in the absence of prior knowledge. By contrast, extreme dealiasing using prior knowledge about silhouettes is demonstrated from a *single* noiseless data frame, where the reconstruction is virtually indistinguishable from the true object. When noise is added to the data, performance degrades gracefully. Even in the presence of significant noise (CNR = 10), the Fourier representation appears qualitatively similar to that of the true object and the recovered silhouette closely resembles the crisp silhouette.

B. Superresolution

In this article, we are invoking the classical meaning of the term superresolution: the recovery of Fourier information beyond the cutoff frequency (or bandwidth extrapolation) through the use of prior knowledge. It is well known that superresolution based on a support constraint (knowledge that the object is of finite extent) can be accomplished, but that it is extremely ill-conditioned [19,20]. Superresolution can also be accomplished with the use of a nonnegativity constraint. The use of superresolution with an opacity constraint has been proposed as well [11]. We demonstrate that dramatic and robust superresolution can be accomplished by utilizing prior knowledge about silhouettes.

Consider the case of a monochromatic point illuminator that casts a crisp shadow onto a surface that is then imaged onto a detector array via an aberration-free lens system. In this case the PSF will be an Airy pattern and the MTF will have a clearly defined cutoff frequency arising from diffraction. Accordingly, the data can be sampled at the Nyquist rate so that there is no aliasing and all Fourier components out to the cutoff frequency can in principle be recovered. It is natural to inquire whether information beyond the cutoff frequency can also be recovered. In our simulation to demonstrate superresolution, we approximated a crisp facial silhouette with a 256×512 representation grid. Figure 14 illustrates that the estimated silhouette successfully recovers Fourier components out to spatial frequencies in excess of eight times the cutoff frequency in the noiseless case. This is a dramatic amount of superresolution. The silhouette estimated from the noiseless sampled data is virtually indistinguishable from the true (crisp) silhouette. Even when a significant amount of noise (CNR = 10) is added to the sampled data, silhouette estimation degrades gracefully. The prior knowledge that we are invoking for estimating facial silhouettes leads to far more robust estimation than superresolution using a support constraint, which is known to be highly ill-conditioned.

6. SUMMARY AND RECOMMENDATIONS

We have demonstrated that fine-resolution estimates of continuously defined facial silhouettes can be obtained from coarsely sampled data in the presence of significant amounts of additive noise. This is accomplished by using strong prior knowledge in the form of a parametric model for facial silhouettes that is far more efficient than a conventional pixel parameterization. Additional prior knowledge obtained through statistical training on an ensemble of facial silhouettes was shown to be useful when the data are particularly

noisy. Practical MAP estimation, accommodating data misregistration and detecting over a finite FOV, was also demonstrated. Finally, both extreme dealiasing and dramatic superresolution (recovery of information beyond the diffraction limit) were demonstrated.

The simulations presented herein are somewhat idealized in that the facial silhouettes are all provided in profile. Such silhouettes arise only in restrictive geometries such that a human face is pointed in a direction orthogonal to the illumination direction and the detector array or surface onto which the silhouette is projected is also orthogonal to the illumination direction. We concentrated on this restrictive case because it retains the essence of the problem of interest and it corresponds to the available training data. An additional projective operator can be included in the data-acquisition model to accommodate projections onto surfaces not perpendicular to the direction of illumination. Note that for extended (e.g., solar) illumination, such geometries will also result in field-dependent blur. This is evident in Fig. 1(b), in which the shadow is relatively crisp where the range from the occluder is small (tire region) and grows progressively more blurry as the range grows (most noticeable at the head). If the geometry of the illumination, occluder, and surface is known, then the field-dependent blur can be readily modeled. Accordingly, the data-agreement term of the MAP estimator will involve a field-dependent blur kernel in the integral instead of a simple convolution kernel.

In the case of using a camera to image a shadow falling on a flat surface, the surface reflectivity may not be uniform. This case can be accommodated by collecting a second image of the surface in the absence of shadow. Similarly, silhouette estimation in the case of a relatively dark object occluding a bright background will require images with and without occlusion when the background is not uniform.

In our experiments, we used a simple spline parameterization to characterize facial silhouettes. We believe that the characterization can be improved upon by using more sophisticated spline parameterizations, such as with NURBS [21] or T-splines [22]. In many silhouette-estimation applications, the parametric model for the silhouette will not be as well behaved as that of the human face in profile. What parametric model should we use to represent the silhouette of a walker for gait recognition or for the silhouette of a satellite for pose estimation? Other candidate parametric models include level sets [23–25] and various shape-coding models [26,27]. Perhaps the parametric silhouette model could be built from the data itself, by adaptively inserting spline control points until the data no longer support further model complexity. The joint estimation of the parametric model and the corresponding parameter values is an important research direction for silhouette estimation and related problems.

ACKNOWLEDGMENTS

Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office.

REFERENCES

1. N. Davidenko, "Silhouetted face profiles: a new methodology for face perception research," J. Vis. **7**(4):6, 1–17 (2007).
2. A. Laurentini, "The visual hull concept for silhouette-based image understanding," IEEE Trans. Pattern Anal. Mach. Intell **16**, 150–162 (1994).
3. E. Boyer and J. S. Franco, "A hybrid approach for computing visual hulls of complex objects," Comput. Vis. Pattern Recogn. **1**, 695–701 (2003).
4. B. DeCann and A. Ross, "Gait curves for human recognition, backpack detection, and silhouette correction in a nighttime environment," Proc. SPIE **7667**, 76670Q (2010).
5. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).
6. C. L. Matson, "Fourier spectrum extrapolation and enhancement using support constraints," IEEE Trans. Signal Process **42**, 156–163 (1994).
7. R. G. Paxman, P. D. Walker, and N. Davidenko, "Silhouette restoration," in *Computational Optical Sensing and Imaging*, OSA Technical Digest (Optical Society of America, 2013), paper CM2C.4.
8. R. G. Paxman, J. H. Seldin, J. R. Fienup, and J. C. Marron, "Use of an opacity constraint in three-dimensional imaging," Proc. SPIE **2241–14**, 116–126 (1994).
9. R. G. Paxman, J. R. Fienup, M. J. Reiley, and B. J. Thelen, "Phase retrieval with an opacity constraint in laser imaging (PROCLAIM)," in *Signal Recovery and Synthesis V*, Technical Digest Series, 11 (Optical Society of America, 1998), pp. 34–36.
10. J. R. Fienup, R. G. Paxman, M. F. Reiley, and B. J. Thelen, "3-D imaging correlography and coherent image reconstruction," Proc. SPIE **3815–07**, 60–69 (1999).
11. R. G. Paxman, "Superresolution with an opacity constraint," in *Signal Recovery and Synthesis III*, Technical Digest Series, Vol. **15** (Optical Society of America, 1989).
12. P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1090–1104 (2000).
13. P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," Image Vis. Comput. **16**, 295–306 (1998).
14. H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley, 2004).
15. A. Berk, L. S. Bernstein, and D. C. Robertson, "MODTRAN: a moderate resolution model for LOWTRAN7," Technical report GL-TR-89-0122 (Air Force Geophysics Laboratory, 1989).
16. R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," SIAM J. Sci. Comput. **16**, 1190–1208 (1995).
17. R. C. Veltkamp, "Shape matching: similarity measures and algorithms," in *Proceedings of International Conference on Shape Modeling and Applications* (IEEE, 2001), pp. 188–197.
18. F. O'Sullivan and G. Wahba, "A cross validated Bayesian retrieval algorithm for nonlinear remote sensing experiments," J. Comput. Phys. **59**, 441–455 (1985).
19. J. L. C. Sanz and T. S. Huang, "Discrete and continuous band-limited signal extrapolation," IEEE Trans. Acoust. Speech Signal Process **31**, 1276–1285 (1983).
20. M. E. Davison, "The ill-conditioned nature of the limited angle tomography problem," SIAM J. Appl. Math. **43**, 428–448 (1983).
21. L. Piegl and W. Tiller, *The NURBS Book* (Springer-Verlag, 1997).
22. T. W. Sederberg, D. L. Cardon, G. T. Finnigan, N. S. North, J. Zheng, and T. Lyche, "T-spline simplification and local refinement," ACM Trans. Graph **23**, 276–283 (2004).
23. T. J. A. Sethian, *Level Set Methods* (Cambridge University, 1996).
24. S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces* (Springer-Verlag, 2002).
25. K. Krishnamurthy, W. U. Bajwa, and R. Willet, "Level set estimation from projection measurements: performance guarantees and fast computation," SIAM J. Imaging Sci. **6**, 2047–2074 4(2013).
26. G. M. Schuster, G. Melnikov, and A. K. Katsaggelos, "Operationally optimal vertex-based shape coding," IEEE Signal Process. Mag.. **15**(6), 91–108 (1998).
27. A. K. Katsaggelos, L. P. Kondi, F. W. Meier, J. Ostermann, and G. M. Schuster, "MPEG-4 and rate-distortion-based shape-coding techniques," Proc. IEEE **86**, 1126–1154 (1998).