

Texture Enhanced Appearance Models

Rasmus Larsen ^{a,*} Mikkel B. Stegmann ^a Sune Darkner ^a
Søren Forchhammer ^b Timothy F. Cootes ^c
Bjarne Kjær Ersbøll ^a

^a*Informatics and Mathematical Modelling, Technical University of Denmark,
Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark*

^b*Research Centre COM, Technical University of Denmark,
Building 345v, DK-2800 Kgs. Lyngby, Denmark*

^c*Division of Imaging Science and Biomedical Engineering,
University of Manchester, Manchester, UK*

Abstract

Statistical region-based registration methods such as the Active Appearance Model (AAM) are used for establishing dense correspondences in images. At low resolution, images correspondences can be recovered reliably in real-time. However, as resolution increases this becomes infeasible due to excessive storage and computational requirements. We propose to reduce the dimensionality of the textural components by selecting a subset of basis functions from a larger dictionary, estimate regression splines and model only the coefficients of the retained basis functions. We demonstrate the use of two types of bases, namely wavelets and wedgelets. The former extends the previous work of Wolstenholme and Taylor where Haar wavelet coefficient subsets were applied. The latter introduces the wedgelet regression tree based on triangulated domains. The wavelet and wedgelet regression splines are functional descriptions of the intensity information and serve to 1) reduce noise and 2) produce a compact textural description. Dimensionality reduction by subsampling in the CDF 9-7 wavelet and wedgelet representations yield better results than 'standard' subsampling in the pixel domain. We show that the bi-orthogonal CDF 9-7 wavelet yields better results than the Haar wavelet. Further, we show that the inherent frequency separation in wavelets allows for cost-free band-pass filtering, e.g. edge-emphasis, and that this edge enhancement provide better results in terms of segmentation accuracy. Wedgelet representation are superior to wavelet representations at high dimensionality-reduction rates. At low reduction rates an edge enhanced wavelet representation provides better segmentation accuracy than the full standard AAM model.

Key words: registration, dimensionality reduction, atlases, deformable models, active appearance models, wavelets, wedgelets, face images

1 Introduction

Since its introduction the Active Appearance Model (AAM) framework [6, 8] has been applied successfully to registration of many types of deformable objects in images (e.g. faces, cardiac ventricles, brain structures [8, 15, 16, 19]). It is based on the estimation of linear models of shape and texture variation by the use of principal components analysis of landmark coordinates and pixel intensities and subsequent inference of model parameters from unseen images by a tangent plane approximation of a shape compensated image manifold.

Modelling every pixel intensity is manageable for low-resolution 2D images. But moving to high-resolution 2D images, 3D and even 3D time-series, this approach is rendered at best very slow and at worst infeasible due to excessive storage and computational requirements.

In order to overcome this problem various alternatives to modelling the raw pixel intensities have been considered. Cootes et al. [7] used a subsampling scheme to reduce the texture model by a ratio of 1:4. The scheme selected a subset of the pixel intensities based on the ability of each pixel to predict corrections of the model parameters. When exploring different multi-band appearance representations Stegmann and Larsen [20] studied the registration accuracy of facial AAMs at different scales in the range $10^3 - 10^5$ pixels obtained by pixel averaging.

In this paper we will take the path of using linear functional descriptions of the underlying intensity patterns and carrying out truncated principal component decomposition of the parameter set of the functionals in order to extract a texture model. These parameter sets will typically be of much lower dimensionality than the number of pixels in the images. In particular, we will use linear functionals based on wavelet and wedgelet basis representation of the texture. Wavelets as well as wedgelets are able to represent piecewise continuous functions. This is an important property when modelling real world images. Wolstenholme and Taylor [21] incorporated a truncated Haar wavelet representation into the AAM framework and evaluated it on a brain MRI data set at a reduction of the number of coefficients of 1:20.

Donoho [10] suggested the wedgelet representation for image texture as a means of edge detection and image compression. An image is represented by a collection of dyadically organized indicator functions with a variety of locations, scales and orientations. The wedgelet tree is a quadtree [11] with terminal nodes being either a dyadic element (degenerate wedgelet) or an

* Corresponding author

Email address: r1@imm.dtu.dk (Rasmus Larsen).

URL: www.imm.dtu.dk/~aam/ (Rasmus Larsen).

affinely split dyadic element (non-degenerate wedgelet). The classification and regression tree (CART) algorithm [4] uses sequential binary splitting of the spatial domain parallel to the coordinate axes, with splits allowed at every data point. In contrast to this, the wedgelet regression trees obey special constraints: only dyadic partitioning (i.e. recursive midpoint splitting) is allowed, with the additional feature that at each terminal node a set of affine splits are also applicable. The constrained splitting leads to fast algorithms. Within each resulting image terminal node (wedge or square) the pixel values are regressed to their mean value.

We generalize the wedgelet transform to triangulated domains (cf. triangulated quadtrees [3]). This has the major advantage of rendering the wedgelet representation independent of piece-wise affine warps of the triangulated domain. Such piece-wise affine warps are customarily chosen in AAM for their speed [19] and the triangulated wedgelet representation thus embraces this choice. The wedgelet transform results in a truncated change of basis for the texture and is represented by a regression tree. The regression tree is estimated using the minimization of the cross validation prediction error across the training set.

The registration accuracy in wavelet and wedgelet based AAMs is evaluated for a case of human face registration using cross validation.

2 Active Appearance Models

AAMs establish a compact parametrization of object variability as learned from a training set by estimating a set of latent variables. The modelled object properties are usually shape and pixel intensities. The latter is hence forward denoted texture. By exploiting prior knowledge of the nature of the optimization space, these models of shape and texture can be rapidly fitted to unseen images, thus providing image interpretation through synthesis.

Training examples are defined by marking up each example image with points of correspondence (i.e. landmarks) over the set either by hand, or by semi-to completely automated methods. From these landmarks a shape model [9] is built. Further, given a suitable warp function a dense (i.e., per-pixel) correspondence is established between the convex hull of the landmarks in each training example. Thus we allow for modelling of texture variability.

Joint variability in shape and texture is modelled by a set of truncated principal components, estimated by an eigen-analysis of the dispersions of shape and texture across the training set. The shape examples are aligned to a common mean using a Generalized Procrustes Analysis (GPA) [2, 13] where all

effects of translation, rotation and scaling are removed. The obtained Procrustes shape coordinates are subsequently projected onto the tangent plane of the shape manifold at the mean shape. The texture examples are warped into correspondence using a piece-wise affine warp and subsequently sampled from this shape-free reference. Typically, this geometrical reference shape is the Procrustes mean shape.

Let $\mathbf{s}_i = \text{vec}\{(x_{ijk})\}$, $i = 1, \dots, I$, $j = 1, \dots, J$, $k = 1, \dots, K$ be J landmarks in K dimensions sampled from a training set of I images, and let $\mathbf{t}_i = \text{vec}\{(y_{ilm})\}$, $l = 1, \dots, L$, $m = 1, \dots, M$ be pixel intensities sampled at L sites in M color components for the same I training images. Furthermore, let $\bar{\mathbf{s}}$ and $\bar{\mathbf{t}}$ denote the mean shape and texture. Synthesized examples are parameterized by $\boldsymbol{\theta}$ and generated by

$$E\{\mathbf{s}\} = \bar{\mathbf{s}} + \boldsymbol{\Phi}_s \boldsymbol{\theta} \quad (1)$$

$$E\{\mathbf{t}\} = \bar{\mathbf{t}} + \boldsymbol{\Phi}_t \boldsymbol{\theta} \quad (2)$$

where $\boldsymbol{\Phi}_s$ and $\boldsymbol{\Phi}_t$ contain the first p eigenvectors of the estimated joint dispersion matrix of the shape and texture vectors, \mathbf{s}_i and \mathbf{t}_i . Equations (1) and (2) constitute the appearance model.

In addition to the parameters $\boldsymbol{\theta}$ the four scalar parameters of a 2D Euclidean similarity group are also needed. These four parameters accounting for scale, orientation, and translation are denoted $\boldsymbol{\psi}$. In order to infer the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ of a previously unseen image a Gaussian error model between model and pixel intensities is assumed. Furthermore, a linear relation between changes in parameters and difference between model and image pixel intensities $\Delta \mathbf{t}$ is assumed, i.e.

$$\Delta \mathbf{t} = \mathbf{X} \begin{bmatrix} \Delta \boldsymbol{\psi} \\ \Delta \boldsymbol{\theta} \end{bmatrix}. \quad (3)$$

\mathbf{X} may be estimated by weighted averaging over perturbations of model parameters and training examples. For an in depth description of AAM and the software implementation used the reader is referred to [8, 19], respectively.

The relation in Eq. (3) is inverted using the least squares solution

$$\begin{bmatrix} \widehat{\Delta \boldsymbol{\psi}} \\ \widehat{\Delta \boldsymbol{\theta}} \end{bmatrix} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Delta \mathbf{t} = \mathbf{Q} \Delta \mathbf{t}. \quad (4)$$

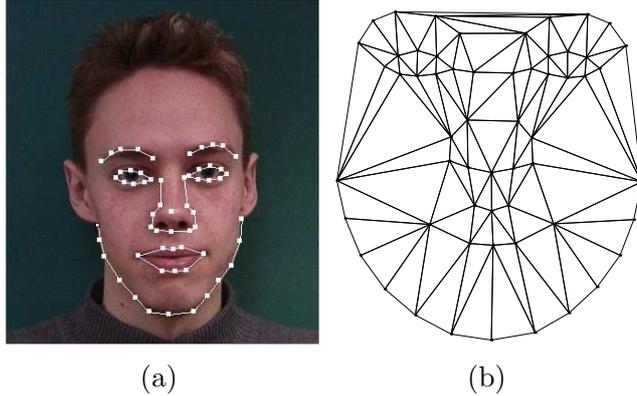


Fig. 1. (a) Human face annotated with 58 landmarks; (b) triangulated mean shape.

The computational problem lies in the repeated application of this relation in the innermost loop of the fitting algorithm. \mathbf{Q} is a non-sparse matrix of dimensions $(p+4) \times (LM)$. LM increases exponentially with numbers of spatial dimension. To reduce the computational burden we propose to use a truncated basis for the representation of the pixel intensities. This introduces the added overhead of transforming between image pixel intensities and this new representation. However, if this transform is based on a sparse matrix, as is the case with wavelet and wedgelet transforms, the computational burden can be considerably reduced.

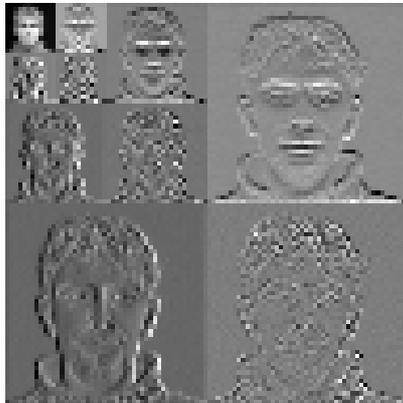
3 Wavelet Enhanced Appearance Modelling

Wavelets are a family of basis functions that decompose signals into both space and frequency. In the following we use the discrete wavelet transform [14], which can be viewed as a set of linear, rank-preserving matrix operations. In practice these are carried out in a convolution scheme known as the fast wavelet transform (FWT) [1] where an image is decomposed by a high-pass filter into sets of detail wavelet sub-bands, and by a low-pass filter into a scaling sub-band. These bands are then down-sampled and can be further decomposed. We use the dyadic (octave) decomposition scheme that successively decomposes the scaling sub-band, yielding a discrete frequency decomposition. Alternative decomposition schemes include the wavelet packet basis where successive decompositions are carried out in the detail sub-bands as well.

Figure 2 shows a three-level octave wavelet decomposition. The first, third and fourth quadrants are the detail sub-bands and stem from the initial decomposition (level 1). The first, third and fourth sub-quadrants of the second quadrant are detail sub-bands from the second decomposition (level 2). Finally, the sub-sub-quadrants of the second sub-quadrant are detail sub-bands from the third decomposition (level 3) with the scaling sub-band at the top



(a)



(b)

Fig. 2. (a) Face image; (b) the wavelet coefficients of a three-level octave decomposition using the Haar wavelet.

left corner.

Wavelets are invertible, which is typically achieved by orthogonality. Wavelet transforms can thus be considered a rotation in function space, which – through successive decompositions – adds a notion of scale. This scale property lends itself nicely to progressive signal processing. Wavelets for image compression are designed to perform rotations that decorrelate image data by using vanishing moments. Wavelet coefficients close to zero can thus be removed with minimal impact on the reconstruction.

Bi-orthogonal wavelets (cf. [5]) will also be considered in the following. These are not strictly orthogonal and therefore in contrast to orthogonal wavelets have odd numbers of filter taps and linear phase filters. They come in pairs of analysis and synthesis filters, together forming a unitary operation.

Now we introduce a notation for wavelet representation and describes how it can be integrated into an AAM framework thereby obtaining a wavelet enhanced appearance model.

First, let an n -level wavelet transform be denoted by

$$\mathcal{W}(\mathbf{t}) = \mathbf{\Gamma}\mathbf{t} = \hat{\mathbf{w}} = [\hat{\mathbf{a}}^\top \hat{\mathbf{u}}_1^\top \cdots \hat{\mathbf{u}}_n^\top]^\top \quad (5)$$

where $\hat{\mathbf{a}}$ and $\hat{\mathbf{u}}$ denote scaling and detail wavelet coefficients, respectively. For 2D images each set of detail coefficients is an ensemble of horizontal, vertical and diagonal filter responses. Reduction of dimensionality is now obtained by a truncation of the wavelet coefficients

$$\mathcal{C}(\hat{\mathbf{w}}) = \mathbf{C}\hat{\mathbf{w}} = \mathbf{w} = [\mathbf{a}^\top \mathbf{u}_1^\top \cdots \mathbf{u}_n^\top]^\top \quad (6)$$

where \mathbf{C} is a modified identity matrix, with rows corresponding to truncated coefficients removed.

As in [21] a wavelet enhanced appearance model is built on the truncated wavelet basis, $\mathbf{w} = \mathcal{C}(\mathcal{W}(\mathbf{t}))$, rather than the raw image intensities in \mathbf{t} . This splits all texture-related matrices into scale-portions. For the texture PCA of wavelet coefficients we have

$$\mathbf{w} = \bar{\mathbf{w}} + \mathbf{\Phi}_w \mathbf{b}_w \Leftrightarrow \quad (7)$$

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{a}} \\ \bar{\mathbf{u}}_1 \\ \vdots \\ \bar{\mathbf{u}}_n \end{bmatrix} + \begin{bmatrix} \mathbf{\Phi}_a \\ \mathbf{\Phi}_{u_1} \\ \vdots \\ \mathbf{\Phi}_{u_n} \end{bmatrix} \mathbf{b}_w$$

where $\mathbf{\Phi}_w$ is the eigenvectors of the wavelet coefficient covariance matrix. Rearranging this into scaling and detail terms we get

$$\mathbf{a} = \bar{\mathbf{a}} + \mathbf{\Phi}_a \mathbf{b}_w \quad (8)$$

and

$$\{\mathbf{u}_i = \bar{\mathbf{u}}_i + \mathbf{\Phi}_{u_i} \mathbf{b}_w\}_{i=1}^n. \quad (9)$$

The texture model is thus inherently multi-scale and may be used for analysis/synthesis at any given scale. Motivations for doing so include robustness and computational efficiency. Compared to multi-scale AAMs this also gives a major decrease in storage requirements.

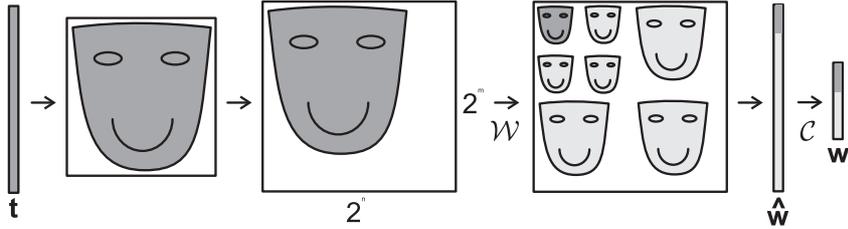


Fig. 3. Two-level wavelet decomposition of a texture vector: the texture vector \mathbf{t} is represented in image space; the image space is padded to have dimensions that are powers of 2; the n -level discrete wavelet transform (\mathcal{W}) is applied. The darker shaded coefficients represent scaling (or low frequency) components, the brighter shaded coefficients represent detail (or high frequency) components; all wavelet coefficients are concatenated to form a wavelet coefficient vector $\hat{\mathbf{w}}$; $\hat{\mathbf{w}}$ is pruned to form the final truncated wavelet representation of the texture vector \mathbf{w} - usually most scaling coefficients are retained.

Using non-truncated orthogonal wavelets (i.e. $\mathbf{C} = \mathbf{I}$), $\mathcal{W}(\mathbf{t})$ is a true rotation in texture hyperspace. Hence the wavelet PCA is a rotation of the original intensity PCA, i.e. $\Phi_{\mathbf{w}} = \Gamma \Phi_{\mathbf{t}}$, iff. \mathcal{W} is fixed over the training set. The PC scores are identical, $\mathbf{b}_{\mathbf{w}} = \mathbf{b}_{\mathbf{t}}$. If \mathbf{C} is chosen to truncate the wavelet basis along directions with near-zero magnitude, wavelet PC scores obviously resemble the original PC scores closely.

Direct usage of the sparse matrix Γ is excessively slow. Instead the fast wavelet transform (FWT) is applied. Figure 3 shows the stages of transformation. First, a normalized texture vector is rendered into its equivalent shape-free image. Secondly, the shape-free image is expanded into a dyadic image representation to avoid any constraints on n due to image size. This image is then transformed using FWT and rendered into the vector $\hat{\mathbf{w}}$ by masking out areas outside the octave representation of the reference shape. Finally, $\hat{\mathbf{w}}$ is truncated into \mathbf{w} .

3.1 Free Parameters and Boundary Effects

To apply the above, the type of wavelet, \mathcal{W} , must be chosen and values for n and \mathbf{C} must be determined. The key issue of estimating \mathbf{C} is described in Section 3.3.

The choice of wavelet type depends on three factors, i) the nature of the data, ii) image-boundary issues, and iii) computational issues. Data containing sharp edges suggests sharp filters such as the Haar wavelet and smooth data requires smooth wavelets.

Because the wavelets operate in a finite discrete domain, boundary issues arise. To calculate filter responses for a full image, boundaries are typically extended

by mirroring pixel values across the boundaries. The width of the boundary extension is half the width of the wavelet filter. Normally, this is carried out as a rectangular extension but in this application the border extension adapts to the shape of the texture image.

Finally, the Haar wavelet represents the computationally simplest wavelet filter having only two non-zero coefficients. In contrast the CDF 9-7 wavelet has 9 non-zero coefficients in the encoding filter and 7 non-zero coefficients in the decoding filter.

3.2 Model Building

Building a wavelet enhanced appearance model can be summarized in five major steps:

- (1) Sample all training textures into $\{\mathbf{t}_i\}_{i=1}^I$.
- (2) Transform $\{\mathbf{t}_i\}_{i=1}^I$ into $\{\hat{\mathbf{w}}_i\}_{i=1}^I$.
- (3) Estimate $\hat{\mathbf{w}}$ and \mathbf{C} .
- (4) Truncate $\{\hat{\mathbf{w}}_i\}_{i=1}^I$ into $\{\mathbf{w}_i\}_{i=1}^I$.
- (5) Build an AAM on $\{\mathbf{w}_i\}_{i=1}^I$.

Further, all incoming textures in subsequent optimization stages should be replaced with their truncated wavelet equivalents, i.e. $\mathcal{C}(\mathcal{W}(\mathbf{t}))$. The synthesis of a wavelet enhanced appearance model is the reverse of Fig. 3, again with appropriate masking and mirroring. Truncated wavelet coefficients are reconstructed using the corresponding mean values:

$$\hat{\mathbf{w}}_{\text{synth}} = \mathcal{C}_{\text{synth}}(\mathbf{w}) = \mathbf{C}^T \mathbf{w} + \mathbf{C}_{\text{synth}} \overline{\hat{\mathbf{w}}} \quad (10)$$

where $\mathbf{C}_{\text{synth}}$ is a modified identity matrix, with rows corresponding to non-truncated coefficients zeroed.

3.3 Wavelet Coefficient Selection

For registration purposes the optimal \mathbf{C} is given by the minimum average error between the optimized model shape and the ground truth shape over a test set of t examples

$$\arg \min_{\mathbf{C}} \left(\sum_{i=1}^t |\mathbf{s}_{i,\text{model}} - \mathbf{s}_{i,\text{ground truth}}|^2 \right), \quad (11)$$

subject to the constraint that \mathbf{C} has g rows. This gives rise to a dimensionality reduction of ratio $1 : L/g$.

However, direct optimization of Eq. (11) is not feasible since each cost function evaluation involves building a complete model from the training set and a subsequent evaluation on a test set. The traditional approach when dealing with reduction of training data ensembles, also taken by [21], is to let \mathbf{C} preserve per-pixel variance over the training set.

First, we compute the empirical variance of each coefficient across the training set, i.e.

$$\hat{\sigma}_j^2 = \frac{1}{I} \sum_{i=1}^I (w_{ij} - \bar{w}_{.j})^2, \quad (12)$$

where w_{ij} denotes the j th wavelet coefficient of the i th training image, $\bar{w}_{.j}$ is the empirical mean of these j th coefficients across training images. Second, we construct \mathbf{C} to select the g coefficients that express the largest variance $\hat{\sigma}_j^2$.

In order to regularize this selection procedure we carry out a smoothing of the wavelet coefficients in the image domain, i.e. intra-quadrant smoothing of the wavelet representation as shown in Fig. 2(b).

3.4 Signal Manipulation

An important added bonus from the wavelet representation is that certain signal manipulations become exceedingly simple. Due to the band separation, frequency response modifications are easily carried out. The simplest modification is to change the norm of the high- and low-pass filters used. To preserve variance the filters must be normalized. With a minor abuse of notation this is denoted by $\|\mathcal{W}\|_2 = 1$. We call this the *normalized case*. To emphasize high-frequency content $\|\mathcal{W}\|_2$ must be less than one. In the following we propose to use the norm $\|\mathcal{W}\|_2 = 1/\sqrt{2}$ and call this the *edge-enhanced case*.

4 Wedgelet Enhanced Appearance Modelling

The wedgelet approach is a way of representing images locally, orientation adaptively and at the appropriate scale. It involves a very simple basis used at different scales. The formulation in [10] is for a square dyadic domain but the nature of the AAM imposes a shift of basis from dyadic to triangulated domains.

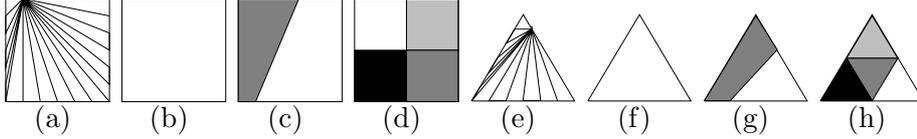


Fig. 4. Templates on the square dyadic domain and the triangulated domain. Plots (a) and (e) show the applicable affine splits beginning at a particular perimeter point.

For each node in the wedgelet tree we may consider the following three wedgelets types:

- (1) a degenerate wedgelet, this is a terminal node without an affine split, cf. Figs. 4(b) and 4(f);
- (2) a non-degenerate wedgelet, this is a terminal node with an affine split, cf. Figs. 4(c) and 4(g), in Figs. 4(a) and 4(e) all affine splits beginning at a single border point are shown;
- (3) an interior node corresponding to a step through scale space, cf. Figs. 4(d) and 4(h).

The wedgelet decomposition is seen to be embedded into a quadtree structure. In this structure the templates are the nodes and the step through scale space the branches. Furthermore, the terminal nodes are all either degenerate or non-degenerate wedgelets. The resulting structure is a regression tree [4]. The corresponding regression model predicts the pixel intensity – greyscale, RGB or other – at the l th image coordinate \mathbf{x}_l in the i th image, y_{il} , with a constant μ_{ir} in each region r for each image

$$f_i(\mathbf{x}_l) = \sum_r \mu_{ir} I\{\mathbf{x}_l \in r\}. \quad (13)$$

$I\{\mathbf{x}_l \in r\}$ is an indicator function returning 1 if \mathbf{x}_l belongs to r and 0 otherwise. Note that for notational simplicity we have dropped the index m for color component on the intensity values, y_{il} . For $M > 1$, y_{il} is a vector of intensities of the M color components modelled. For the sum of squared error loss criterion $\sum_l \|y_{il} - f_i(\mathbf{x}_l)\|^2$ it is easy to see that the optimal μ_{ir} is just the average of y_{il} in region r , $\hat{\mu}_{ir} = \text{ave}(y_{il} | \mathbf{x}_l \in r)$.

4.1 Optimal Partitioning

The optimal partitioning is found by a bottom-up approach. For each triangle at each level we seek the model that minimizes the C -fold cross-validation estimate of the prediction error across all affine splits/no split. Specifically,

let $\kappa : \{1, \dots, I\} \mapsto \{1, \dots, C\}$ be an indexing function that indicates the partition to which training object (image) $i = 1, \dots, I$ is allocated by randomization, and denote by $\hat{s}^{-\kappa(i)}$ a split estimated with the $\kappa(i)$ 'th part removed. $\hat{a} = a(\hat{s}^{-\kappa(i)})$ and $\hat{b} = b(\hat{s}^{-\kappa(i)})$ are the regions resulting from splitting a dyadic square or triangle by this affine split $\hat{s}^{-\kappa(i)}$, and \hat{c} is the entire dyadic square or triangle. Furthermore, let the regression parameters from the i th image resulting from applying the split $\hat{s}^{-\kappa(i)}$ be

$$\hat{\mu}_{ir} = \text{ave}(y_{il} | \mathbf{x}_l \in r), \quad r \in \{\hat{a}, \hat{b}, \hat{c}\}. \quad (14)$$

Then the cross-validation errors become

$$\begin{aligned} CVE_{split} &= \sum_{i=1}^I \left[\sum_{\mathbf{x}_l \in \hat{a}} \|y_{il} - \hat{\mu}_{i\hat{a}}\|^2 + \sum_{\mathbf{x}_l \in \hat{b}} \|y_{il} - \hat{\mu}_{i\hat{b}}\|^2 \right] \\ CVE_{no\ split} &= \sum_{i=1}^I \sum_{\mathbf{x}_l \in \hat{c}} \|y_{il} - \hat{\mu}_{i\hat{c}}\|^2 \end{aligned} \quad (15)$$

The optimal split/no split cross-validation errors for a triangle and its three siblings are then compared to those for their parent in order to determine if a non-degenerate or a degenerate wedgelet should be declared, or if the four siblings should be merged into a triangle or a dyadic at the next higher (parent) level.

In order to be able to control the dimensionality reduction ratio obtained, we add a complexity penalty to the error term over which we carry out cross-validation. This complexity penalty is proportional to the image variance, σ^2 and the cardinality of the wedgelet tree. λ^2 is the proportionality constant, i.e.

$$CP(\lambda) = \lambda^2 \cdot \sigma^2 \text{card}(\mathcal{P}). \quad (16)$$

\mathcal{P} is the partition, and $\text{card}(\mathcal{P})$ is the cardinality of \mathcal{P} . The multiplication of the CP with the image variance compensates for (gain) differences between training images. Except for the image variance weight this is the same complexity penalty proposed by Donoho [10] for the case of wedgelet compression over a single image.

The optimization over all affine splits is conducted as an exhaustive search over a discretization (cf. Figs. 4(a) and 4(e)) corresponding to the pixel size. The

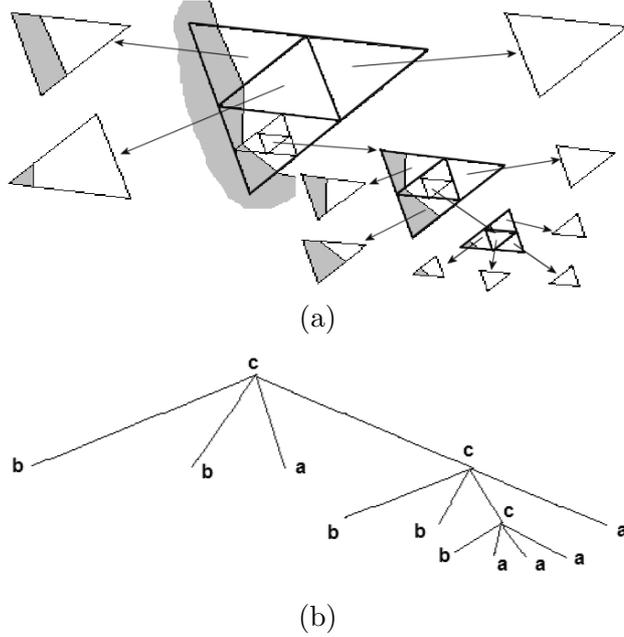


Fig. 5. A representation of (a) a binary image and (b) the resulting tree structure; the tree nodes are of types *a* degenerate, *b* non-degenerate, *c* interior.

indexing of pixels within a triangle and computation of areas are conveniently done by the use of barycentric coordinates [17]. Figure 5 shows how a result on a binary image might look.

4.2 Model building

When working on an AAM, an initial triangulation is available from the annotation and Delaunay triangulation on the mean shape (cf. Fig. 1(b)). This initial collection of triangles will be the root of the tree. From here each branch will be equivalent to the type of tree shown in Fig. 5.

After having grown a common wedgelet tree as described in the previous section for the training set, we can proceed to train the wedgelet enhanced appearance model. As before the shape is described by the landmark coordinates, i.e. $\mathbf{s}_i = \text{vec}\{(x_{ijk})\}$. However, for the texture we substitute the wedgelet coefficients for the original intensity samples, i.e. we use $\boldsymbol{\mu}_i = \text{vec}\{(\hat{\mu}_{irm})\}$, $r = 1, \dots, R$, $m = 1, \dots, M$. R is the number of regions used by the wedgelet tree.

4.3 Computational Requirements

The active part of the wedgelet enhanced appearance model is trained using Eqs. (3) and (4). The resulting \mathbf{Q} matrix has dimensions $(p+4) \times (RM)$. In each iteration of model to image fitting this is also the number of multiplications and additions to be carried out. However, we must also take into consideration the added overhead of calculating the wedgelet representation of the image patch that is covered by the model in each iteration. Letting every pixel fall into just one region of the wedgelet tree the computational load in computing the regional means is essentially LM additions and RM multiplications. Let the dimensionality reduction ratio of the wedgelet representation be $\eta = L/R$, then the reduction in computational load is better than a factor

$$\frac{RM + (p+4)RM}{(p+4)LM} = \left[\frac{1}{p+4} + 1 \right] \frac{1}{\eta} \approx \frac{1}{\eta} \quad (17)$$

As we shall see dimensionality reduction ratios of 1:100 are achievable. We have here ignored the computational load related to warping the image patch to the model. This is the same whether we use wedgelets or not and is conveniently and very quickly accomplished using modern graphics hardware [19].

5 Results and Discussion

Data for the experiments are an image database of 37 annotated faces. Each image is a 640×480 RGB image of a face of an adult human. The data set consists of images of 7 female and 30 male faces. Each face has been manually annotated with 58 corresponding landmarks (see [19] for a detailed analysis). The average landmark distance from model to ground truth (pt.pt.) was used as performance measure. Model searches were initialized by displacing the mean configuration $\pm 10\%$ of its width and height in x and y from the optimal position. Cross validation experiments were carried out to assess the performance of the proposed algorithms.

5.1 Wavelet enhancement

Two wavelet bases were evaluated; the orthogonal Haar wavelet and the widely used bi-orthogonal CDF 9-7 wavelet (cf. [5]). Both of these were tested in the normalized and the edge-enhanced case using dimensionality reduction ratios

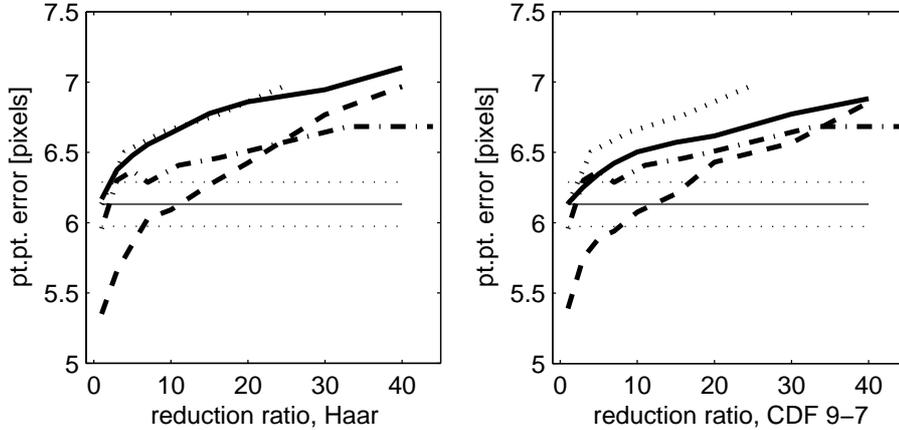
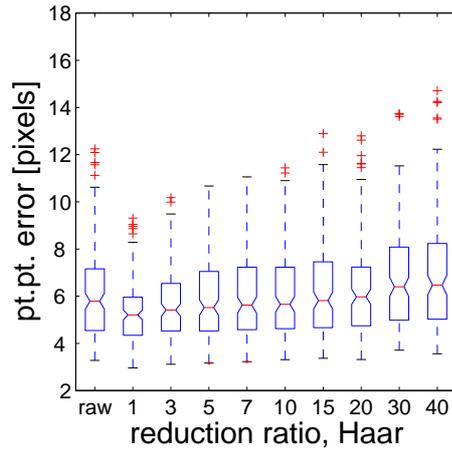


Fig. 6. The average registration error vs. the dimensionality reduction ratio for the truncated wavelet and wedgelet representations of the texture. We show the registration error for the $\|\mathcal{W}\|_2 = 1$ wavelet selection weighting scheme (solid), the $\|\mathcal{W}\|_2 = 1/\sqrt{2}$ wavelet selection weighting scheme (dashed), standard subsampling AAM (dotted), and wedgelet AAM (dash-dotted). In the left plot we have applied the Haar wavelet and in the right plot the CDF 9-7 wavelet. The standard subsampling AAM and wedgelet AAM are the same for the two plots. The horizontal lines show values for a full standard AAM. The standard error is 0.15

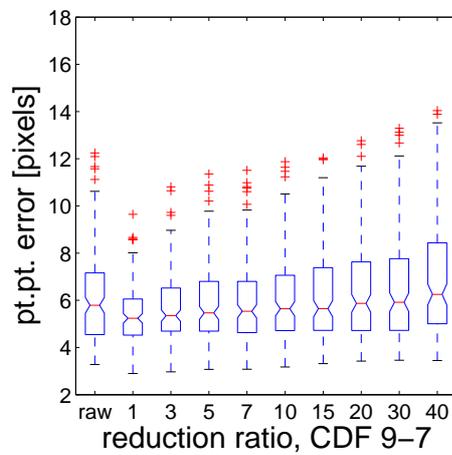
in the range 1:1 – 1:40. We compare the wavelet enhanced methods to the standard uncompressed AAM as well as a series of experiments where standard AAM are built from subsampled images as described in [7].

All experiments used three wavelet decomposition levels. Results are shown in Figs. 6 and 7. The standard AAM contained 31224 pixels on average over all leave-one-out experiments.

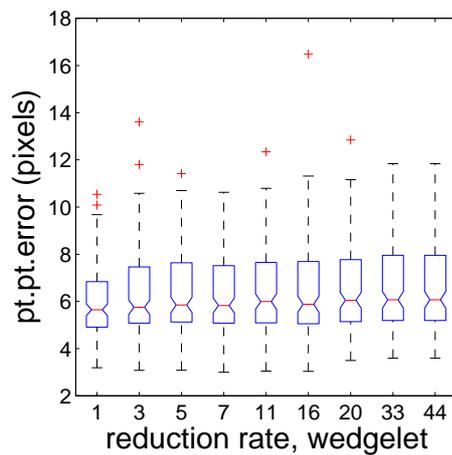
Our first observation from Fig. 6 is that the average segmentation accuracy degrades gracefully with increasing dimensionality reduction ratio. Furthermore, the CDF 9-7 wavelet outperforms the Haar wavelet representation for all reduction ratios. The results for the Haar wavelet using the normalized weighting scheme coincides with the results obtained for simple subsampling



(a)



(b)

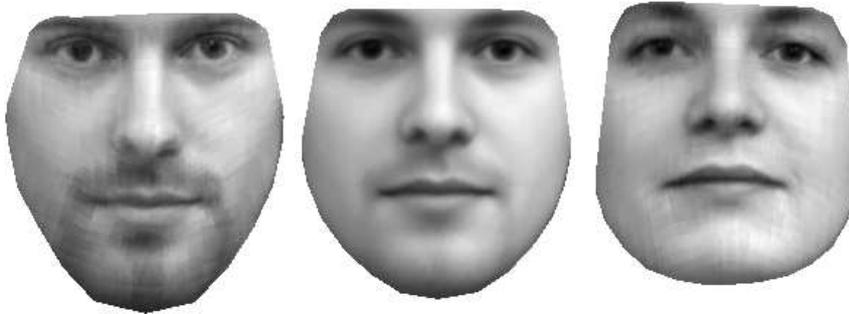


(c)

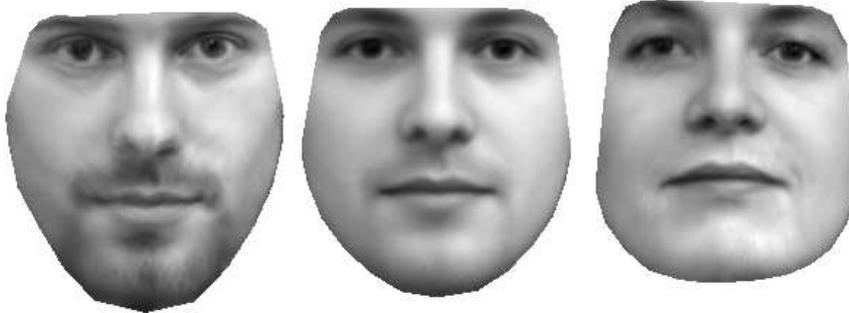
Fig. 7. Registration accuracy determined by cross validation. The abscissae are the ratio between the number of pixels and the number of retained wavelet coefficients or wedges in the model. This is the dimensionality reduction ratio. The ordinate is the mean point to point prediction error evaluated at the original 58 landmarks for all faces in the cross validation study.



Fig. 8. Selected wavelet coefficients for the face training set (CDF, ratio 1:10). The scaling coefficients are shown in the upper left. Left: $\|\mathcal{W}\|_2 = 1$. Right: $\|\mathcal{W}\|_2 = 1/\sqrt{2}$.



(a) Haar - PC_1



(b) CDF 9-7 - PC_1

Fig. 9. The first combined mode of texture and shape variation; $c_1 = \{-3\sigma_1, 0, 3\sigma_1\}$. (a) wavelet enhanced appearance model (b) Haar, ratio 1:10; (b) CDF 9-7, ratio 1:10.

of pixels in the texture representation. The CDF 9-7 wavelet representation with the normalized weighting scheme performs better than pixel-wise sub-sampling.

However, the most striking result is that putting an increased weight on the high frequency wavelet bands by using the weighting scheme $\|\mathcal{W}\|_2 = 1/\sqrt{2}$

(i.e. enhancing the edges) results in a much better performance for all reduction ratios. We even observe for both wavelets that for reduction ratios between 1 and 10 the segmentation accuracy is better than for the full un-reduced standard AAM model. In conclusion pure intensity based AAM methods are outperformed by methods that put an emphasis on edge information.

In addition to the overall average performance of the methods we also compare the error distribution over test examples for the wedgelet AAM and for both wavelet AAMs using weighting scheme $\|\mathcal{W}\|_2 = 1/\sqrt{2}$. In Fig. 7 standard boxplots of the mean error for all test faces are shown for all dimensionality reduction ratios. The plots should be interpreted as follows: each box has horizontal lines at the lower quartile, median and upper quartile of the distribution. The lines extending from each box show the range of the rest of the data. Single crosses beyond the end of these lines are outliers with values above 1.5 times the interquartile range from the respective lower or upper quartile. The notches at the median (the rotated V-shapes at the sides of each box) are robust estimates of the uncertainty about the medians for box-to-box comparison. If the notches of pairs of boxes do not overlap, then the true medians are significantly different at the 0.05 significance level by Wilcoxon’s rank sum test [12].

First off all this confirms that the “raw” standard AAM is significantly outperformed wrt. segmentation accuracy by the edge-enhanced wavelet AAM applying low dimensionality reduction rates. Furthermore, we can conclude that no significant decrease in segmentation accuracy is seen, as we move from no dimensionality reduction to a dimensionality reduction by a factor of 40. We do however, see a slight increase in variance of the errors for increased dimensionality reduction rates for the wavelet cases.

In order to investigate the edge enhancing weighting scheme further, Fig. 8 shows the selected wavelet coefficients using the CDF 9-7 wavelet at ratio 1:10 using both normalized and edge-enhancing filters. The normalized case tends to preserve mostly low-frequency content, while the edge-enhanced case distributes wavelets coefficients near image details more evenly across levels. Together with the nature of the weighting this leads to models with more emphasis on edges.

Finally, Fig. 9 shows the first combined mode of texture and shape deformation for two compressed models at dimensionality reduction ratio 1:10. Subtle blocking artifacts from the Haar wavelet are present, while the smooth 9-7 leaves a more visually pleasing synthesis result. Quantitatively, the 9-7 wavelet also offered a better reconstruction of the training textures in terms of mean squared error (MSE).

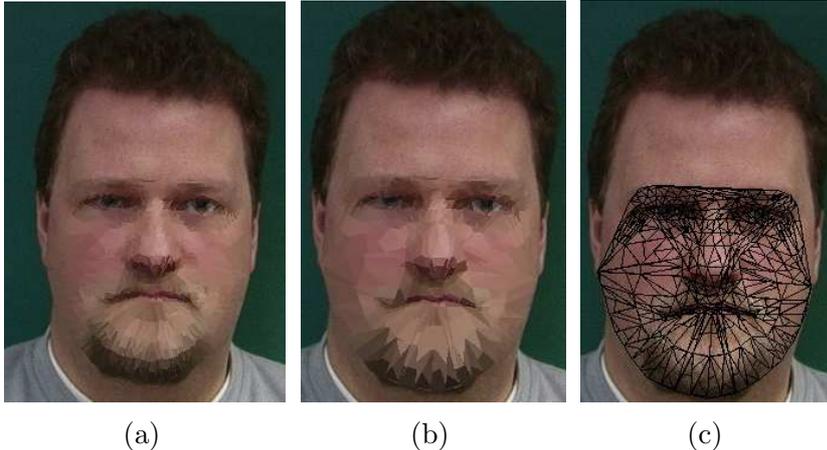


Fig. 10. Images compressed using triangulated wedgelets. (a) 1:3 ratio and (b) 1:40 using the triangulation from Fig. 1(b). (c) Result (b) superimposed with the subdivided mesh.

5.2 Wedgelet enhancement

Figure 10 shows the result of compressing a single image using the approach described above. The original Delauney triangulation of the face data set (cf. Fig. 1(b)) is subdivided using the penalized complexity criterion in Eqs. (15) and 16) above. In Fig. 10 wedgelet dimensionality reduction at ratios 1:3 and 1:40 for a single face are shown.

In Fig. 11 the combined principal components of the texture descriptors μ_{irm} and the tangent space aligned landmark coordinates are shown. Comparing Figs. 11(a) with a dimensionality reduction ratio of 1:3 and Figs. 11(b) with a dimensionality reduction ratio of 1:40, we see that the first principal components contain the same variations independently of the dimensionality reduction ratios. This leads us to conjecture that the wedgelet representation similarly to the wavelet case indeed dismisses irrelevant noise components and retains original signal information.

In Fig. 12 examples of the segmented facial features using a wedgelet enhanced AAM with dimensionality reduction ratios 1:3 and 1:40 are shown. Again the results are indistinguishable.

In order to compare the registration quality of the wedgelet enhanced AAM and its ability to perform with increasing dimensionality reduction ratio we have conducted a cross-validation study across the construction of the wedgelet decomposition of the training set and the construction of the AAM. The average point to point landmark distance from model to ground truth has been used to measure the performance. The models were initialized using a displacement of $\pm 10\%$ of the width and the height of the mean AAM model from the optimal position in the x and y direction.

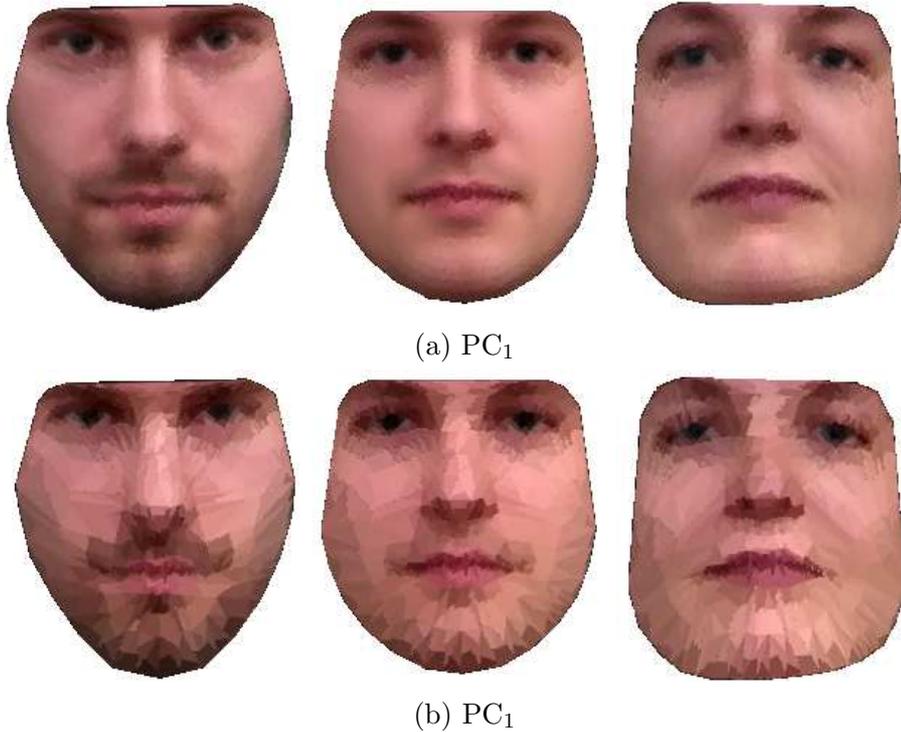


Fig. 11. 1st principal components shown at +3 standard deviation, mean and -3 standard deviation for (a) wedgelet dimensionality reduction ratio 1:3, (b) wedgelet dimensionality reduction ratio 1:40.

For face registration we show dimensionality reduction rates of 1:150 with a decrease in registration accuracy of 8%. By comparison, for brain registration the wavelet compressed AAMs reports dimensionality reduction rates of 1:20 with a decrease in registration accuracy of 7% [21].

As expected a slight decrease in performance is seen as the dimensionality reduction ratio increases (cf. Figs. 6 and 7(c)). However, using wedgelets yields good results all the way up to 1:150, (cf. Fig. 13). A major difference between wedgelets and wavelets is in the synthesized image. The truncated wavelet representation yields nice and smooth synthetic images very pleasing to the eye. However, it should be noted that the purpose of the wedgelet enhanced appearance model is registration with minimum storage and computational cost and not image reconstruction. Therefore the model should not be evaluated on the "blockyness" of Figs. 10 and 11.

Fig. 10 serves to demonstrate where in the images important information regarding registration is present. Fig. 11 demonstrates that the low order principal components of the uncompressed and the compressed data set are similar.

The truncated wavelet transform is computationally more demanding than the truncated wedgelet transform for the same number of basis functions. Initial off-line determination of the wedgelet regression tree is computationally

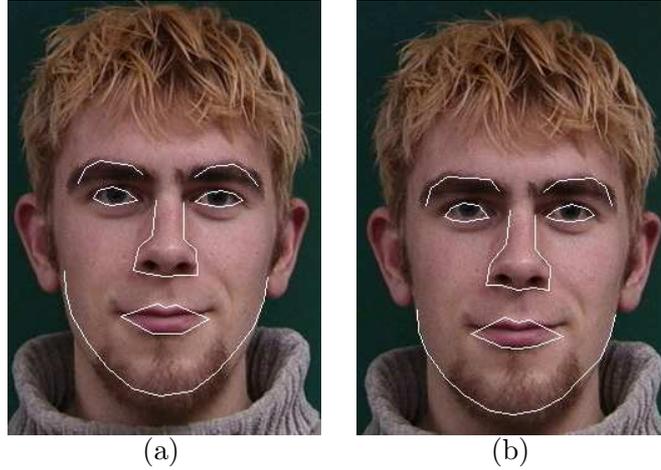


Fig. 12. (a) Wedgelet AAM using a wedgelet dimensionality reduction rate of 1:3, (b) 1:40.

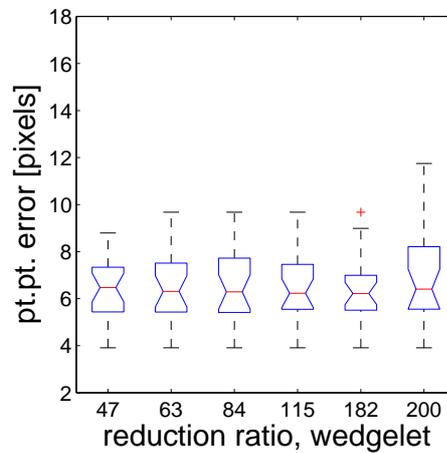


Fig. 13. Additional wedgelet registration accuracy study determined by cross validation for high reduction ratios. The abscissa is the ratio between the number of pixels and the number of wedges in the model. The ordinate is the mean point to point prediction error evaluated at the original 58 landmarks for all faces in the cross validation study.

expensive. However, wedgelets reduce the texture descriptor size and hereby reduce the computational cost and storage cost significantly due to the reduction of \mathbf{Q} in Eq. (4).

6 Conclusion

This paper extends the previous work of Wolstenholme and Taylor where an AAM was augmented with the Haar wavelet and evaluated on brain MRI using a fixed dimensionality reduction ratio of 1:20. Our work validates their

previous findings in a different case study using a more thorough evaluation methodology. In addition, the more recent CDF 9-7 wavelet is evaluated and compared to the Haar wavelet. The CDF 9-7 wavelet proves to yield better accuracy for the same reduction rate than the Haar wavelet representation.

We have also shown that the inherent frequency separation in wavelets allows for simple band-pass filtering, which enables dimensionality reduction schemes that both decrease complexity *and* increase registration accuracy. Increased registration accuracy is seen for Haar as well as CDF 9-7 wavelet representations. In particular, we have shown that applying a weighting scheme that emphasize high frequency content results in an improved segmentation accuracy over the standard intensity based AAM.

Although no significant difference was observed between the two wavelets in terms of registration accuracy; the CDF 9-7 wavelet performed consistently better. Where the Haar wavelet did not perform better than standard pixel subsampling this was the case for the CDF 9-7 wavelet. Further, the CDF 9-7 also provided the best synthesis quality. However, if low computational complexity is required the Haar wavelet should be chosen.

Our case study supports the expected behavior that it is the high-frequency image content that provides the registration accuracy, i.e. edges hold precise information about position. Variance preserving filters are optimal for reconstruction but not necessarily for registration. Inherent scale-separation has been obtained in the presented unified AAM and DWT framework and enabled a very simple method for favoring high-frequency content. This is obtained while still retaining the robustness gained from a low frequency representation of the object.

Furthermore, we have defined a 2D wedgelet transform on the triangulated domain. We have used cross-validation to arrive at a truncated wedgelet representation of the texture in an active appearance model setting. The triangulated wedgelet transform embraces the triangulated domains used in AAMs.

The original wedgelet representation uses a complexity penalized (CP) RSS criterion. For the multi-object (image) situation we introduce a new ensemble complexity penalized cross-validation error. This complexity penalty favors relative larger wedges/triangles while compensating for scaling differences/warps and gain differences between images.

In applying the wedgelet transform to ensembles of face images we arrive at dimensionality reduction rates up to 1:150 with only subtle degradation of the registration accuracy. The truncated wedgelet representation yielded better segmentation accuracy than the wavelet representations for the same reduction ratios.

We have used the wavelet and wedgelet based functional representation of the image intensity patterns in order to obtain a more parsimonious description of the texture. However, we can also think of such a functional representation as a way of regularizing the principal components solution (cf. [18]).

7 Acknowledgements

We gratefully acknowledge the support from The Danish Medical Research Council, grant no. 52-00-0767, and The Danish Technical Research Council, grants no. 26-01-0198 and no. 2059-03-0032. Furthermore, DTU students M. M. Nordstrøm, M. Larsen, and J. Sierakowski are thanked for collecting and annotating the face database used in this article. 9

References

- [1] M. Antonini, M. Barlaud, P. Mathieu, I. Daubechies, Image coding using wavelet transform, *IEEE Transactions on Image Processing* 1 (2) (1992) 205–220.
- [2] J. M. F. ten Berge, Orthogonal Procrustes rotation for two or more matrices, *Psychometrika* 42 (1977) 267–276.
- [3] M. W. Bern, D. Eppstein, J. R. Gilbert, Provably good mesh generation, in: *Proc. 31st Symp. Foundations of Computer Science, Vol. I, IEEE, 1990*, pp. 231–241.
- [4] L. Breiman, J. Friedman, R. Olshen, C. Stone, *Classification and regression trees*, Wadsworth & Brooks/Cole advanced books & software, Monterey, California, 1984, 358 pp.
- [5] A. Cohen, I. Daubechies, J.-C. Feauveau, Biorthogonal bases of compactly supported wavelets, *Comm. Pure and Applied Mathematics* 45 (1992) 485–560.
- [6] T. F. Cootes, G. J. Edwards, C. J. Taylor, Active appearance models, in: H. Burkhardt, B. Neumann (Eds.), *Proceedings of the European Conference On Computer Vision (ECCV), Freiburg, Germany, June 2-6, Vol. 1406 of Lecture Notes in Computer Science, Springer, Heidelberg, Germany, 1998*, pp. I-484–I-498.
- [7] T. F. Cootes, G. Edwards, C. J. Taylor, A comparative evaluation of active appearance model algorithms, in: *BMVC 98. Proc. of the Ninth British Machine Vision Conf., Vol. 2, Univ. Southampton, 1998*, pp. 680–689.
- [8] T. F. Cootes, G. J. Edwards, C. J. Taylor, Active appearance models, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (6) (2001) 681–685.

- [9] T. F. Cootes, C. J. Taylor, D. H. Cooper, J. Graham, Active shape models – their training and application, *Computer Vision, Graphics and Image Processing* 61 (1) (1995) 38–59.
- [10] D. Donoho, Wedgelets: Nearly minimax estimation of edges, *Annals of Statistics* 27 (1999) 859–897.
- [11] R. A. Finkel, J. L. Bentley, Quad trees: A data structure for retrieval on composite keys, *Acta Informatica* 4 (1974) 1–9.
- [12] J. D. Gibbons, *Nonparametric Statistical Inference*, 2nd Edition, M. Dekker, 1985.
- [13] J. C. Gower, Generalized Procrustes analysis, *Psychometrika* 40 (1975) 33–50.
- [14] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation., *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 11 (7) (1989) 674–693.
- [15] S. Mitchell, B. Lelieveldt, R. Geest, J. Schaap, J. Reiber, M. Sonka, Segmentation of cardiac MR images: An active appearance model approach, in: *Medical Imaging 2000: Image Processing*, San Diego CA, SPIE, Vol. 1, SPIE, 2000.
- [16] S. C. Mitchell, B. P. F. Lelieveldt, R. J. van der Geest, H. G. Bosch, J. H. C. Reiber, M. Sonka, Multistage hybrid active appearance model matching: Segmentation of left and right ventricles in cardiac MR images, *IEEE Transactions on Medical Imaging* 20 (5) (2001) 415–423.
- [17] A. F. Möbius, *Der barycentrische Calcul, ein neues Hilfsmittel zur analytischen Behandlung der Geometrie / dargestellt und angewendet*, Leipzig, Germany, 1827.
- [18] J. O. Ramsay, B. W. Silverman, *Functional Data Analysis*, Springer Series in Statistics, Springer Verlag, New York, 1997.
- [19] M. B. Stegmann, B. K. Ersbøll, R. Larsen, FAME – a flexible appearance modelling environment, *IEEE Transactions on Medical Imaging* 22 (10) (2003) 1319–1331.
- [20] M. B. Stegmann, R. Larsen, Multi-band modelling of appearance, *Image and Vision Computing* 21 (1) (2003) 61–67.
- [21] C. B. H. Wolstenholme, C. J. Taylor, Wavelet compression of active appearance models, in: *Medical Image Computing and Computer-Assisted Intervention, MICCAI, 1999*, pp. 544–554.