



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Computer Vision
and Image
Understanding

Computer Vision and Image Understanding xxx (2004) xxx–xxx

www.elsevier.com/locate/cviu

Force field feature extraction for ear biometrics

David J. Hurley, Mark S. Nixon*, John N. Carter

Department of Electronics and Computer Science, University of Southampton, SO17 1BJ, UK

Received 4 September 2003; accepted 10 November 2004

Abstract

The overall objective in defining feature space is to reduce the dimensionality of the original pattern space, whilst maintaining discriminatory power for classification. To meet this objective in the context of ear biometrics a new force field transformation treats the image as an array of mutually attracting particles that act as the source of a Gaussian force field. Underlying the force field there is a scalar potential energy field, which in the case of an ear takes the form of a smooth surface that resembles a small mountain with a number of peaks joined by ridges. The peaks correspond to potential energy wells and to extend the analogy the ridges correspond to potential energy channels. Since the transform also turns out to be invertible, and since the surface is otherwise smooth, information theory suggests that much of the information is transferred to these features, thus confirming their efficacy. We previously described how field line feature extraction, using an algorithm similar to gradient descent, exploits the directional properties of the force field to automatically locate these channels and wells, which then form the basis of characteristic ear features. We now show how an analysis of the mechanism of this algorithmic approach leads to a closed analytical description based on the divergence of force direction, which reveals that channels and wells are really manifestations of the same phenomenon. We further show that this new operator, with its own distinct advantages, has a striking similarity to the Marr-Hildreth operator, but with the important difference that it is non-linear. As well as addressing faster implementation, invertibility, and brightness sensitivity, the technique is also validated by performing recognition on a database of ears selected from the XM2VTS face database, and by comparing the results with the more established technique of Principal Components Analysis. This confirms not only that ears do indeed

* Corresponding author.

E-mail addresses: djh@analyticalengines.co.uk (D.J. Hurley), msn@ecs.soton.ac.uk (M.S. Nixon), jnc@ecs.soton.ac.uk (J.N. Carter).

appear to have potential as a biometric, but also that the new approach is well suited to their description, being robust especially in the presence of noise, and having the advantage that the ear does not need to be explicitly extracted from the background.

© 2004 Elsevier Inc. All rights reserved.

1. Introduction

The potential of the human ear for personal identification was recognized and advocated as long ago as 1890 by the French criminologist Alphonse Bertillon [1]. In machine vision, ear biometrics has received scant attention compared to the more popular techniques of automatic face, eye, or fingerprint recognition. However, ears have played a significant role in forensic science for many years, especially in the United States, where an ear classification system based on manual measurements has been developed by Iannarelli, and has been in use for more than 40 years [2], although the safety of ear-print evidence has recently been challenged in the Courts [3,4].

Ears have certain advantages over the more established biometrics; as Bertillon pointed out, they have a rich and stable structure that is preserved from birth well into old age. The ear does not suffer from changes in facial expression, and is firmly fixed in the middle of the side of the head so that the immediate background is predictable, whereas face recognition usually requires the face to be captured against a controlled background. Collection does not have an associated hygiene issue, as may be the case with direct contact fingerprint scanning, and is not likely to cause anxiety, as may happen with iris and retina measurements. The ear is large compared with the iris, retina, and fingerprint and therefore is more easily captured.

Burge and Burger [5,6] were amongst the first to describe the ear's potential as a biometric using graph matching techniques on a Voroni diagram of curves extracted from the Canny edge map. Moreno et al. [9] have tackled the problem with some success using neural networks and have reported a recognition rate of 93% using a two-stage neural network technique. Chang et al. [7,8] use principal components analysis, and they also did a comparison of ear and face recognition, where they came to the interesting conclusion that ears are essentially just as good as faces for machine recognition. They have reported a recognition rate of almost 91% using a multi-modal PCA approach. In our original descriptions of our new approach to low-level feature extraction [10,11] we also explored the feasibility of recognizing people by their ears. In addition to extending and improving the basic technique, we now validate it using ear images derived from a standard database. Our results are broadly comparable to the aforementioned results, and when taken together, certainly suggest that ear biometrics has a good future.

Our technique provides a robust and reliable description of the ear without the need for explicit ear extraction. It has two distinct stages: *Image to Force Field Transformation*, and *Force Field Feature Extraction*. First, the entire image is transformed into a force field by pretending that each pixel exerts a force on all the other pixels,

which is proportional to the pixel's intensity and inversely proportional to the square of the distance to each of the other pixels. It turns out that treating the image in this way is equivalent to passing it through an extremely powerful low pass filter which transforms it into a smooth undulating surface, but with the interesting property that the new surface retains all the original information. Operating in the force field domain allows us access to a wealth of established vector calculus techniques to extract information about this surface. Also, because it is based on a natural force field there is the prospect of implementing the transform in silicon hardware by mapping the image to an array of electric charges. The powerful smoothing also affords valuable resistance to noise and surface matching is also greatly facilitated when the surfaces are smooth.

The smooth surface corresponds to the potential energy field underlying the vector force field, and in our earlier work [10,11] we showed how the directional properties of the force field can be exploited to automatically locate a small number of potential wells and channels which correspond to local energy peaks and ridges, respectively, that form the basis of the new features. Information theory would also suggest that these are good features as follows. A smooth surface is easier to describe than a complicated one and therefore requires less information. In the extreme, a hemisphere can be described just by its radius. At the other extreme, a surface made up of random noise would be very hard to describe and therefore would require much more information. The total image information is conserved in the transformation, so we argue that there must be a redistribution of information density away from the smoothest parts towards the salient channels and wells which break up the monotony of the otherwise smooth surface.

In this paper we show how the divergence operator may be used to extend the earlier algorithmic approach leading to an even richer functional form. Others have used force field techniques to good effect in the past. Luo et al. [14] describe the use of Vector Potential to extract corners by treating the Canny edge map of an image as a current density. Ahuja [15] used a novel force field segmentation technique where pixels of similar intensity were detected by assigning inter-pixel forces inversely proportional to the gray level difference. Ahuja and Chuang [16] used a potential field model to extend the medial axis transform. Xu and Prince [17] extended the active contour model by replacing the external local force with a force field derived from the image edge map. Similarly, Yahia et al. [18] used an external force field to attract a level-set towards extracted image-pixels in a deformable structure application. Also Siddiqi et al. [20,21] make extensive use of vector flow techniques in their work.

What makes our application considerably different from the others is that our force field is derived directly from the entire image without any pre-processing whatsoever; we effectively smooth the entire image with a gigantic $1/r$ kernel, which is more than four times the area of the original image, thus obliterating all fine detail such as edges, yet preserving all the image information.

Our target application is ear biometrics where we would expect measurement conditions such as diffuse-lighting and viewpoint to be carefully controlled. We shall assume that the subject will be looking directly ahead, sideways to the camera view,

and the resulting camera projection will be taken as the subject's ear image for biometric purposes, even if the ears should protrude excessively. We do not seek pinna-plane tilt invariance simply because there is no reason why the tilted version of a given pinna should not correspond to another subject. A selection of samples taken from each of 63 subjects drawn from the XM2VTS face profiles database [19] has been used to test the viability of the technique. A classification rate of 99.2% has been achieved so far, using just simple template matching on the basic channel shapes, demonstrating the merit of the technique at least at this scale. It can reasonably be expected that the use of more sophisticated channel matching techniques would lead to even better results. We also do a head-to-head comparison with PCA which also helps to demonstrate the implicit extraction inherent in our technique. As such, a new low-level feature extraction approach has been demonstrated with success in a new application domain.

2. Force field transforms

This section gives a brief description of the mathematical foundations of the new transforms. The basic concepts used can be found in various introductory works on physics [22,23] and electromagnetics [24,25]. We also consider faster computation by applying the convolution theorem in the frequency domain. We also investigate how the features are affected by the combination of the unusual dome shape and changes in image brightness. The question of transform invertibility is considered as this establishes that the transforms are information preserving. Further details of invariance, including initialization invariance, scale invariance, and noise tolerance can be found in [10–13].

The image is transformed to a force field by treating the pixels as an array of mutually attracting particles that act as the source of a Gaussian force field. We use Gaussian force as a generalization of the inverse square laws which govern the gravitational, electrostatic, and magnetostatic force fields, to discourage the notion that any of these forces are in play. The laws governing these forces can all be deduced from *Gauss's Law*, itself a consequence of the inverse square nature of the forces. So, purely as an invention, the pixels are considered to attract each other according to the product of their intensities and inversely to the square of the distances between them. Each pixel is assumed to generate a spherically symmetrical force field so that the total force $\mathbf{F}(\mathbf{r}_j)$ exerted on a pixel of unit intensity at the pixel location with position vector \mathbf{r}_j by a remote pixels with position vector \mathbf{r}_i and pixel intensities $P(\mathbf{r}_i)$ is given by the vector summation.

$$\mathbf{F}(\mathbf{r}_j) = \sum_i \left(\begin{array}{l} P(\mathbf{r}_i) \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} \forall i \neq j \\ 0 \forall i = j \end{array} \right). \quad (1)$$

To calculate the force field for the entire image, this equation should be applied at every pixel position in the image. Units of pixel intensity, force, and distance are arbitrary, as are the co-ordinates of the origin of the vector field.

The defining equations could be applied directly, but in practice for greater efficiency, the process can be treated as a convolution of the image with the force field corresponding to a unit value test pixel, and then invoking the Convolution Theorem to perform the calculation as a multiplication in the frequency domain, the result of which is then transformed back into the spatial domain. The force field equation for an $M \times N$ pixel image becomes,

$$\text{force field} = \sqrt{M \times N} \mathfrak{F}^{-1}[\mathfrak{F}(\text{unit force field}) \times \mathfrak{F}(\text{image})], \quad (2)$$

where \mathfrak{F} stands for the Fourier Transform and \mathfrak{F}^{-1} for its inverse. This applies equally to the energy field to be described next. The usual care must be taken to ensure that dimensions of the unit sample force field are twice those of the image dimensions and that sufficient zero padding is used to avoid aliasing effects. Fig. 1 shows how to implement this in *Mathcad* in which $1j$ denotes the complex operator and **cff** and **icff** denote the Fourier and inverse Fourier transforms, respectively.

There is a scalar potential energy field associated with the vector force field where the two fields are related by the well-known equation [24,25],

$$\mathbf{F}(\mathbf{r}) = -\text{grad}(E(\mathbf{r})) = -\nabla E(\mathbf{r}). \quad (3)$$

This equation tells us that the force at a given point is equal to the additive inverse of the gradient of the potential energy field at that point. This simple relationship allows the force field to be easily calculated by differentiating the energy field and allows some conclusions drawn about one field to be extended to the other.

We can restate the force field formulation in energy terms to derive the energy field equations directly as follows. The image is transformed by treating the pixels as an array of particles that act as the source of a Gaussian potential energy field.

```

ff(pic) :=
  sr ← 2 · (rows(pic) - 1)
  sc ← 2 · (cols(pic) - 1)
  r ← rows(pic) - 1
  c ← cols(pic) - 1
  for rr ∈ 0..sr
    for cc ∈ 0..sc
      upfrr,cc ←  $\frac{(r + c \cdot 1j) - (rr + cc \cdot 1j) + 0j}{(|r + c \cdot 1j - (rr + cc \cdot 1j)|)^3}$ 
  upf3·rows(pic)-3, 3·cols(pic)-3} ← 0
  inp ← pic
  inp3·rows(pic)-3, 3·cols(pic)-3} ← 0
  oup ←  $\sqrt{\text{rows(inp)} \cdot \text{cols(inp)}} \cdot \text{icff}(\overrightarrow{\text{cff}(upf) \cdot \text{cff}(inp)})$ 
  ff ← submatrix(oup, r, 2·r, c, 2·c)

```

Fig. 1. Force field by convolution in Mathcad.

It is assumed that there is a spherically symmetrical potential energy field generated by each pixel, so that $E(\mathbf{r}_j)$ is the total potential energy imparted to a pixel of unit intensity at the pixel location with position vector \mathbf{r}_j by the energy fields of remote pixels with position vectors \mathbf{r}_i and pixel intensities $P(\mathbf{r}_i)$, and is given by the scalar summation

$$E(\mathbf{r}_j) = \sum_i \begin{pmatrix} \frac{P(\mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_j|} \forall i \neq j \\ 0 \forall i = j \end{pmatrix}, \quad (4)$$

where the units of pixel intensity, energy, and distance are arbitrary, as are the coordinates of the origin of the field. To calculate the energy field for the entire image Eq. (4) should be applied at every pixel position. The result of this process for the energy transform for an ear image is shown in Fig. 2, where the same surface has been depicted from a variety of different perspectives below the lobe.

The potential surface undulates, forming local peaks or maxima, with ridges leading into them. These peaks we call potential energy wells since, by way of analogy, if the surface were to be inverted and water poured over it, the peaks would correspond to small pockets where water would collect. Notice that the highest of the three obvious peaks in Fig. 2 has a ridge that slopes gently towards it from the smaller peak to its left. This corresponds to a potential energy channel, because to extend the analogy, water that happened to find its way into its inverted form would gradually flow along the channel towards the peak. The reason for the dome shape of the energy surface can be easily understood by considering the case where the image has just one gray level throughout. In this situation, the energy field at the center would have the greatest share of energy because test pixels at that position would have the shortest average distance between themselves and all the other pixels, whereas test pixels at the corners would have the greatest average distance to all the other pixels, and therefore the least total energy imparted to them.

2.1. An invertible linear transform

The transformation is linear since the energy field is derived purely by summation which is itself a linear operation. What is less obvious is that the transform is also invertible. For an N pixel image, the application of Eq. 4 at each of the N pixel positions leads to a system of N equations in N unknowns. Now if the N equations are

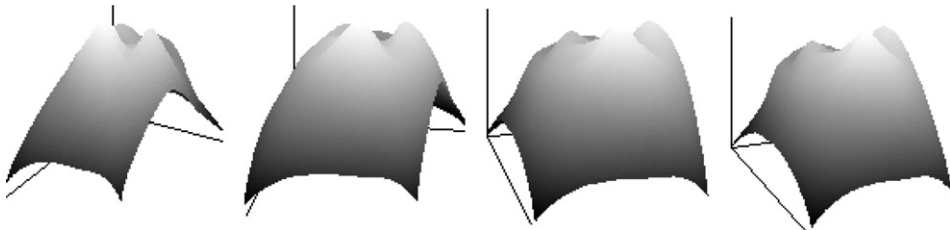


Fig. 2. Energy surface for an ear viewed from below the lobe.

linearly independent, then it follows [26] that the system of equations can be solved for the pixel values, given the energy values. In other words, the transform would be invertible, and the original image could be completely recovered from the energy surface, thus establishing that the transform preserves information. This system of N equations can be expressed as a matrix multiplication of an $N \times 1$ vector of pixel intensities by an $N \times N$ square matrix of coefficients d_{ji} corresponding to the inverse distance scalars given by,

$$d_{ji} = \frac{1}{|\mathbf{r}_j - \mathbf{r}_i|} \quad (5)$$

producing an $N \times 1$ vector of pixel energies. Equation 6 shows this multiplication for a simple 2×2 pixel image.

$$\begin{pmatrix} 0 & d_{01} & d_{02} & d_{03} \\ d_{10} & 0 & d_{12} & d_{13} \\ d_{20} & d_{21} & 0 & d_{23} \\ d_{30} & d_{31} & d_{32} & 0 \end{pmatrix} \begin{pmatrix} P(\mathbf{r}_0) \\ P(\mathbf{r}_1) \\ P(\mathbf{r}_2) \\ P(\mathbf{r}_3) \end{pmatrix} = \begin{pmatrix} E(\mathbf{r}_0) \\ E(\mathbf{r}_1) \\ E(\mathbf{r}_2) \\ E(\mathbf{r}_3) \end{pmatrix}. \quad (6)$$

All the determinants of matrices corresponding to the sequence of square images ranging from 2×2 pixels to 33×33 pixels have been computed and have been found to be non-zero. It has also been verified that all non-square image formats up to 7×8 pixels have associated non-singular matrices [12]. Notwithstanding questions of machine accuracy, these results suggest that the energy transform is indeed invertible for most image sizes and aspect ratios.

2.2. Dome shape and brightness sensitivity

As stated in the introduction, we do not seek viewpoint invariance, or illumination invariance, either in intensity or direction, because we have assumed carefully controlled measurement conditions. However, it is still interesting to investigate how the position of features will be affected by the combination of the unusual dome shape and changes in image brightness. The effect will first be analyzed and then confirmed by experiment. Should the individual pixel intensity be scaled by a factor a and also have an additive intensity component b , we would have,

$$E(\mathbf{r}_j) = \sum_i \left\{ \begin{array}{l} \frac{aP(\mathbf{r}_i)+b}{|\mathbf{r}_i-\mathbf{r}_j|} \forall i \neq j \\ 0 \forall i = j \end{array} \right\} = a \sum_i \left\{ \begin{array}{l} \frac{P(\mathbf{r}_i)}{|\mathbf{r}_i-\mathbf{r}_j|} \forall i \neq j \\ 0 \forall i = j \end{array} \right\} + \sum_i \left\{ \begin{array}{l} \frac{b}{|\mathbf{r}_i-\mathbf{r}_j|} \forall i \neq j \\ 0 \forall i = j \end{array} \right\}. \quad (7)$$

We see that scaling the pixel intensity by the factor a merely scales the energy intensity by the same factor a , whereas adding an offset b is more troublesome, effectively adding a pure dome component corresponding to an image with constant pixel intensity b . This could be corrected by subtracting the dome component, if b can be estimated. The effect of the offset and scaling is shown in Fig. 3 with the channels superimposed. We see that scaling by a factor of 10 in (E) has no effect as expected.

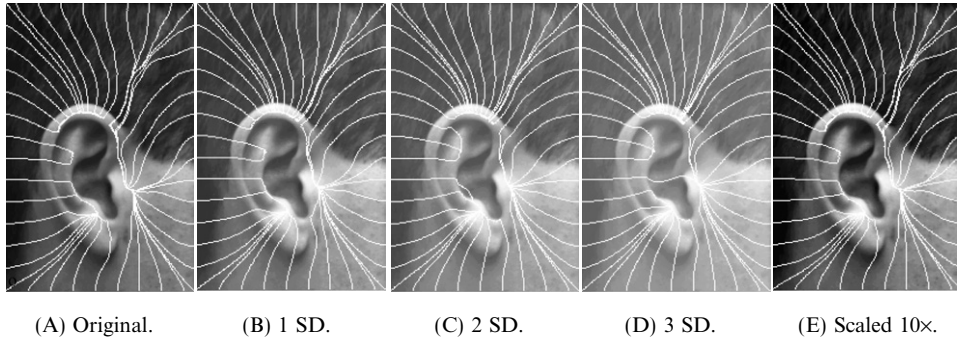


Fig. 3. Effect of additive and multiplicative brightness changes.

The original image in (A) has a mean value of 77 and a standard deviation of 47. Images (B) to (D) show the effect of progressively adding offsets of one standard deviation. At one standard deviation the effect is hardly noticeable and even at 3 standard deviations the change is by no means catastrophic as the channel structure alters little. We therefore conclude that operational lighting variation in a controlled biometrics environment will have little effect. These conclusions are borne out by the results of the corresponding recognition experiments in Table 2.

3. Force field feature extraction

Here field line extraction is presented in outline, as it has already been described in detail in our previous work, whereas the analytic form of convergence feature extraction is presented in detail for the first time. The striking resemblance of convergence to the Marr-Hildreth operator [31] is illustrated and the differences highlighted, especially the non-linearity of the convergence operator. The close correspondence between the field line and convergence techniques is demonstrated by superimposing their results for an ear.

3.1. Field line feature extraction

Fig. 4 demonstrates in brief outline the field line approach to feature extraction for an ear image. A more detailed description can be found in [10–12]. In Fig. 4B a set of 40 test pixels is arranged in an ellipse shaped array around the ear and allowed to follow the field direction so that their trajectories form field lines which capture the general flow of the force field. The test pixel positions are advanced in increments of one pixel width, and test pixel locations are maintained as real numbers, producing a smoother trajectory than if they were constrained to occupy exact pixel grid locations. Notice how ten field lines cross the upper ear rim and how each line joins a common channel that follows the curvature of the ear rim rightwards

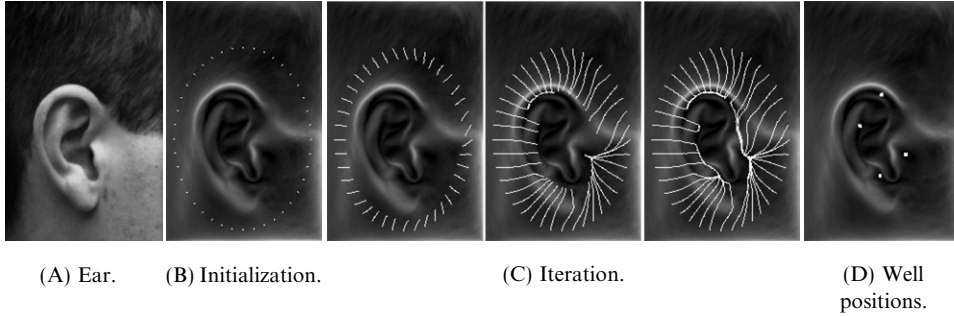


Fig. 4. Field line, channel, and well formation for an ear.

finally terminating in a potential well. The well locations have been extracted by observing clustering of test pixel co-ordinates so that Fig. 4D is simply obtained by plotting the terminal positions of all the co-ordinates.

3.2. Convergence feature extraction

Here we introduce the new analytical method of feature extraction as opposed to the field line method. This method came about as a result of analyzing in detail the mechanism of field line feature extraction. As shown in Fig. 6C, when the arrows usually used to depict a force field are replaced with unit magnitude arrows, thus modeling the directional behavior of exploratory test pixels, it becomes apparent that channels and wells arise as a result of patterns of arrows converging towards each other, at the interfaces between regions of almost uniform force direction. As the divergence operator of vector calculus measures precisely the opposite of this effect, it was natural to investigate the nature of any relationship that might exist between channels and wells and this operator. This resulted not only in the discovery of a close correspondence between the two, but also showed that divergence provided extra information corresponding to the interfaces between diverging arrows.

The concept of the divergence of a vector field will first be explained, and then used to define the new function. Convergence is compared with the Marr-Hildreth operator which is a Laplacian operator and the important difference that convergence is not Laplacian, due to its non-linearity, is illustrated. The function's properties are then analyzed in some detail, and the close correspondence between field line feature extraction and the convergence technique is illustrated by superimposing their results for an ear image.

The divergence of a vector field is a differential operator that produces a scalar field representing the net outward flux density at each point in the field. For the vector force field $\mathbf{F}(\mathbf{r})$ it is defined as,

$$\operatorname{div} \mathbf{F}(\mathbf{r}) = \lim_{\Delta v \rightarrow 0} \frac{\oint \mathbf{F}(\mathbf{r}) \cdot d\mathbf{S}}{\Delta v}, \quad (8)$$

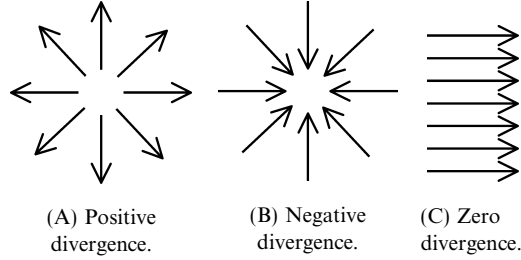


Fig. 5. Divergence of a vector field.

where $d\mathbf{S}$ is the outward normal to a closed surface S enclosing an incremental volume ΔV . In two-dimensional Cartesian co-ordinates it may be expressed as follows [24,25],

$$\text{div}\mathbf{F}(\mathbf{r}) = \nabla \cdot \mathbf{F}(\mathbf{r}) = \left(\frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} \right), \quad (9)$$

where F_x and F_y are the Cartesian components of \mathbf{F} . Fig. 5 illustrates the concept of divergence graphically. In Fig. 5A we see an example of positive divergence where the arrows flow outwards from the center, and in Fig. 5B we see negative divergence, where the arrows flow inwards, whereas in Fig. 5C there is no divergence because all the arrows are parallel.

Having described divergence we may now use it to define convergence feature extraction. Convergence provides a more general description of channels and wells in the form of a mathematical function in which wells and channels are revealed to be peaks and ridges, respectively, in the function value. The new function maps the force field to a scalar field, taking the force as input and returning the additive inverse of the divergence of the force direction. The function will be referred to as the force direction convergence field $C(\mathbf{r})$ or just convergence for brevity. A more formal definition is given by

$$C(\mathbf{r}) = -\text{div}\mathbf{f}(\mathbf{r}) = -\lim_{\Delta A \rightarrow 0} \frac{\oint \mathbf{f}(\mathbf{r}) \cdot d\mathbf{l}}{\Delta A} = -\nabla \cdot \mathbf{f}(\mathbf{r}) = -\left(\frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y} \right), \quad (10)$$

where, $\mathbf{f}(\mathbf{r}) = \frac{\mathbf{F}(\mathbf{r})}{|\mathbf{F}(\mathbf{r})|}$, ΔA is incremental area, and $d\mathbf{l}$ is its boundary outward normal.

This function is real valued and takes negative values as well as positive ones where negative values correspond to force direction divergence. It is interesting to compare this function with the Marr-Hildreth operator given by,

$$\text{MH}(\mathbf{r}) = \text{div}\mathbf{g}(\mathbf{r}) = \nabla \cdot \mathbf{g}(\mathbf{r}) = \left(\frac{\partial g_x}{\partial x} + \frac{\partial g_y}{\partial y} \right), \quad (11)$$

where, $\mathbf{g}(\mathbf{r}) = \text{grad}(G(\mathbf{r}) * I(\mathbf{r}))$, $I(\mathbf{r})$ is the image and $G(\mathbf{r}) = e^{-(r^2/2\sigma^2)}$ is a Gaussian kernel.

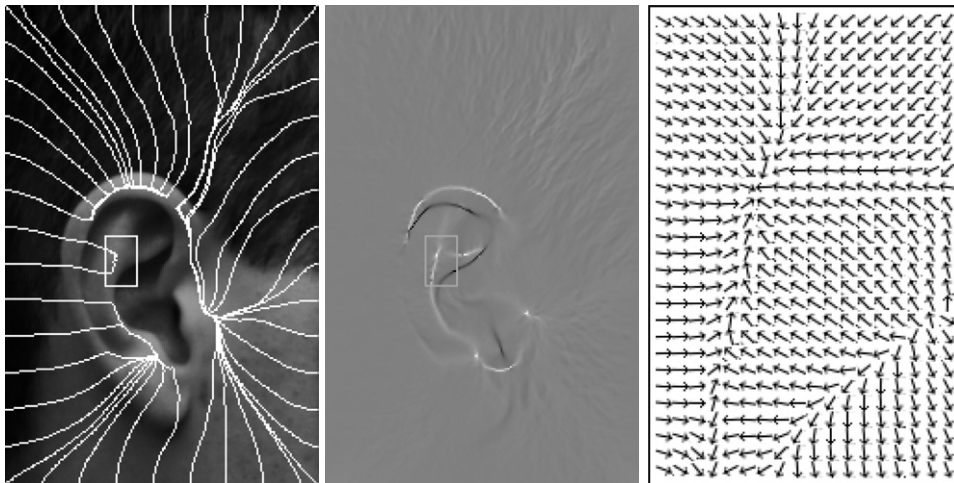
The Marr-Hildreth operator, also known as Laplacian of Gaussian (LoG), uses a Gaussian kernel which is optimal in Gaussian noise, whereas the $1/r$ kernel is an artifact of its force field nature, and the intrinsic smoothing it affords is merely a fortu-

nate consequence of its formulation. We must also stress that whilst Marr-Hildreth is linear, convergence is non-linear because it is based on force direction rather than force. This non-linearity means that we are obliged to perform the operations in the order shown; we cannot take the divergence of the force and then divide by the force magnitude. $\text{Div}(\text{grad}/|\text{grad}|) \neq (\text{div grad})/|\text{grad}|$. This is easily illustrated by a simple example using the scalar field e^x in the following equation:

$$\left\{ \begin{array}{l} \text{div}(\text{grad}/|\text{grad}|) \\ \nabla \cdot \left(\frac{\nabla e^x}{|\nabla e^x|} \right) = \nabla \cdot \frac{e^x \mathbf{i}}{e^x} = \nabla \cdot \mathbf{i} = 0 \end{array} \right\} \neq \left\{ \begin{array}{l} (\text{div grad})/|\text{grad}| \\ \frac{\nabla \cdot \nabla e^x}{|\nabla e^x|} = \frac{e^x}{e^x} = 1 \end{array} \right\}, \quad (12)$$

where \mathbf{i} is a unit vector in the x direction. The convergence is zero because we have a field of parallel unit magnitude vectors, whereas in the second case the vectors are parallel but the magnitude changes, resulting in a net outflow of flux at any point. This illustrates that even though convergence looks very much like a Laplacian operator, it definitely is not.

Fig. 6B shows the convergence field for an ear image, while Fig. 6A shows the corresponding field lines. A magnified version of a small section of the force direction field, depicted by a small rectangular insert in Fig. 6B, is shown in Fig. 6C. In Fig. 6B the convergence values have been adjusted to fall within the range 0–255, so that negative convergence values corresponding to antichannels appear as dark bands, and positive values corresponding to channels appear as white bands. Notice that the antichannels are dominated by the channels, and that the antichannels tend to lie within the confines of the channels. Also, notice how wells appear as bright white spots.

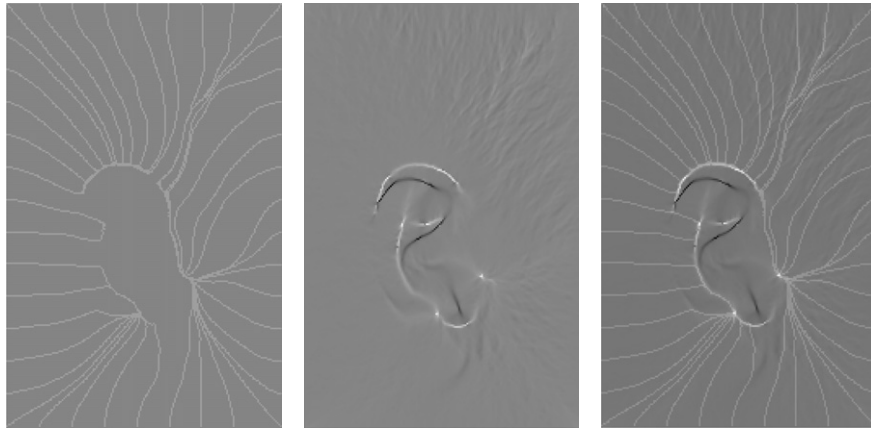


(A) Field line features.

(B) Convergence map.

(C) Magnified insert force field.

Fig. 6. Convergence field.



(A) Field line. (B) Convergence map. (C) Superposition of (A) on (B).

Fig. 7. Correspondence between channels and convergence.

The correspondence between the convergence function and the field line features can be seen by observing the patterns in the force direction field shown in Fig. 6C. Notice the correspondence between the converging arrows and white ridges, and between the diverging arrows and black ridges. It is evident that the convergence map provides more information than the field lines, in the form of negative versions of wells and channels or antiwells and antichannels, although it should be possible to modify the field line technique to extract this extra information by seeding test pixels on a regular grid and reversing the direction of test pixel propagation.

Fig. 7 shows the convergence field of an ear image with the corresponding field lines superimposed. Fig. 7A is the field line map, and Fig. 7B is the convergence map, while Fig. 7C is the superposition of one on the other. We can see clearly how channels coincide with white ridges in the convergence map and that potential wells coincide with the convergence peaks. Notice the extra information in the center of the convergence map that is not in the field line map, illustrating an advantage of convergence over field lines.

4. Ear recognition

The early hope of using just wells as a compact description did not live up to its expectation because some ears produce only two wells. The richer but more complicated description provided by the channels is therefore used instead. The problem of comparing ears has been reduced to one of comparing channels. This should be easier than comparing ears directly but it is still not a trivial task. Fig. 8 shows a synthetic image whose principal field lines have been identified, where these could be defined as the field lines that make the most contribution to a given channel. Assuming that a way can be found of easily extracting such lines they will still need to be compared. Perhaps techniques derived from differential geometry, which deals with

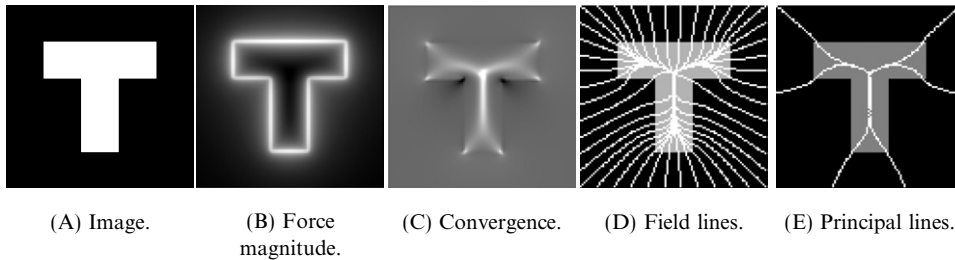


Fig. 8. Principal field lines.

the geometry of curves and surfaces, could be exploited to describe the form of the channels [27]. Perhaps contour description techniques could be used to describe and compare the channels, although it might not be as easy as describing and comparing just simple closed contours [28,29]. Principal field lines would appear to offer a ready made list of points to facilitate contour descriptions, while the convergence approach would require thresholding and thinning to extract the same information.

A lower bound on the validity of the technique is established on a database of 252 ear images by using template matching of thresholded convergence maps, realized using Fourier cross-correlation techniques. The automatic extraction advantage is demonstrated by deliberately not accurately extracting or registering the ears in the sense that the database consists of 141×101 pixel images where the ears have only an average size of 111×73 and are only roughly located by eye in the center of these images. This can be seen clearly in Fig. 10A where we see a marked variation both in vertical and horizontal ear-location. The force field technique gives a correct classification rate of 99.2% on this set, whereas running PCA on the same set gives a result of only 62.4% but when the ears are accurately extracted by cropping to the average ear size of 111×73 , running PCA then gives a result of 98.4%, thus clearly demonstrating the inherent extraction advantage. The effect of brightness change by addition that we described earlier in Fig. 3 is mirrored here by comparing the recognition results with those for PCA when subjected to the same changes, and our earlier conclusion that the technique is robust under variable light conditions is borne out by experiment, where we see that in the worst case for an addition of 3 standard deviations the force field results only change by 2%.

4.1. The ear database

To create a suitable database of experimental ear images the XM2VTS face profiles database [19] was chosen, from which 63 subjects were selected as being suitable candidates for ear recognition by eliminating those whose ears were covered by hair. The ear database thus consists of 4 samples of each of 63 subjects captured during 4 sessions over a 5-month period thus ensuring natural intra-class variation. Diffuse lighting was used during image capture. For each subject there was a significant degree of rotation in the vertical plane containing the line of sight as shown in Fig. 9. We have compensated for this rotation by measuring the angle from the top of the ear to the

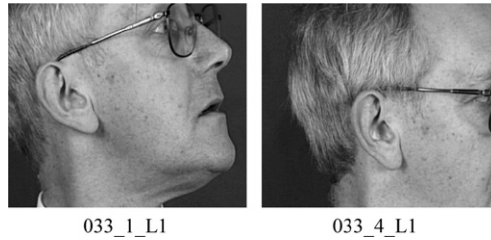


Fig. 9. Difference of up to 20° of rotation in XM2VTS profiles.

corner of the eye and then rotating each of the four images to the mean of the spread. The worst case rotation difference varies between 1.5° and 20° with an average value of about 6° . The standard deviations for each subject vary from 0.6° to 7.3° with an average value of 2.3° . The image rotation function used bilinear interpolation.

Fig. 10 shows how the images of one of the subjects, with the XM2VTS identification number 000, have been processed to extract the convergence maps.

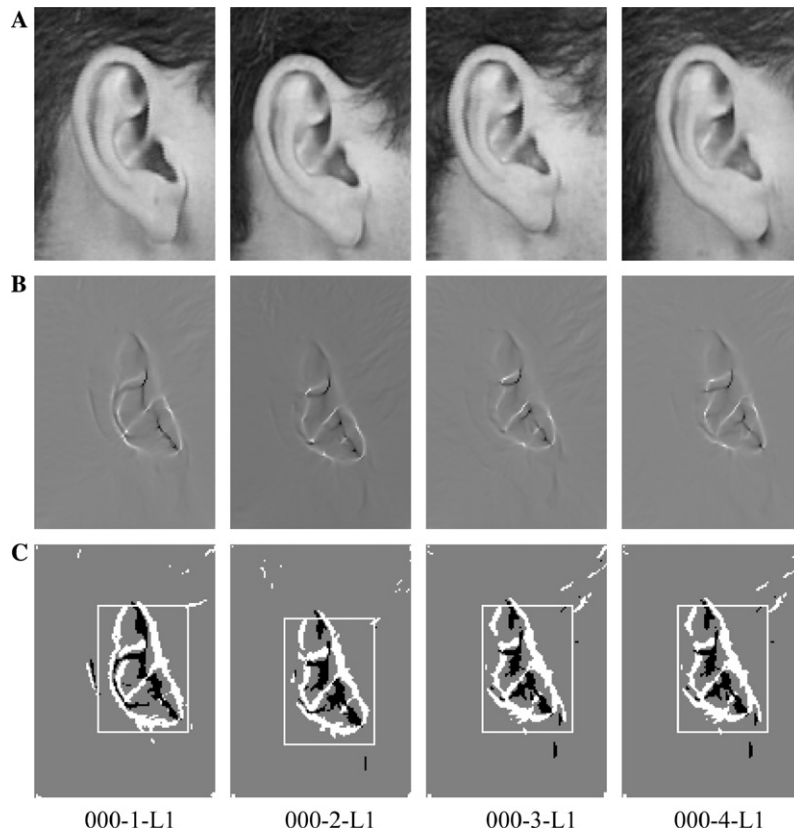


Fig. 10. Feature extraction for subject 000. (A) 141×101 Ear images. (B) Convergence fields. (C) Thresholded convergence maps.

The original color face profiles were converted to gray scale and a 141×101 pixel frame was manually adjusted by eye to surround and crop the ear images (A). The force field transforms of the ear images were taken and the force fields were converted to convergence fields (B).

4.2. Template matching

Fourier based cross-correlation techniques are used to perform multiplicative template matching on ternary thresholded convergence maps. Levels less than minus one standard deviation are mapped to -1 , whilst those greater than one standard deviation map to $+1$ and those remaining map to 0 . Thresholding is done not only in terms of the convergence level but also in terms of a rectangular exclusion zone centered on the convergence magnitude centroid shown in Fig. 10C. The centroid of the convergence tends to be stable with respect to the ear features and this approach has the advantage of removing unwanted outliers such as bright spots caused by spectacles. The threshold of one standard deviation was chosen experimentally resulting in the template channel thickness shown in Fig. 10C. The size of the rectangle was chosen as 71×51 pixels by adjusting its proportions to give a good fit for most of the convergence maps. Notice how for image 000-2 which is slightly lower than the other three, that the centroid-centered rectangle has correctly tracked the template downwards.

Fig. 11 shows a selection of convergence maps for the first eight subjects. These results also confirm the stability of the centroids with respect to the features, since visual inspection of the results confirm that all the templates are correctly aligned in the middle of their frames. If a particular centroid were unstable, then the corresponding template would appear offset.

4.3. Classification results

The thresholded convergence maps were exhaustively cross-correlated to yield a 252×252 element array of cross-correlations. The matrix of correlations is too large to be displayed directly, but can be illustrated graphically as a gray-scale image and as a confusion matrix, as shown in Fig. 12. The self-correlations along the main diagonal are excluded, as indicated by a black line along the diagonal. The position corresponding to the highest value in each row is assigned a white pixel while all the others are set to black. As expected most of the white pixels lie within a 4×4 pixel square centered on the main diagonal, indicating correct classification. The two solitary white pixels shown lying outside these squares indicate incorrect classifications. Out of the 252 trials, 250 resulted in correct classification which corresponds to a classification rate of 99.2%.

Fig. 13 shows the frequency distributions for the intra-class and inter-class correlations. The close fit of a normal distribution, corresponding to their respective means and standard deviations, suggests that they might indeed be normally distributed. The cross-correlation axes have been divided into 25 intervals, between minimum and maximum values.

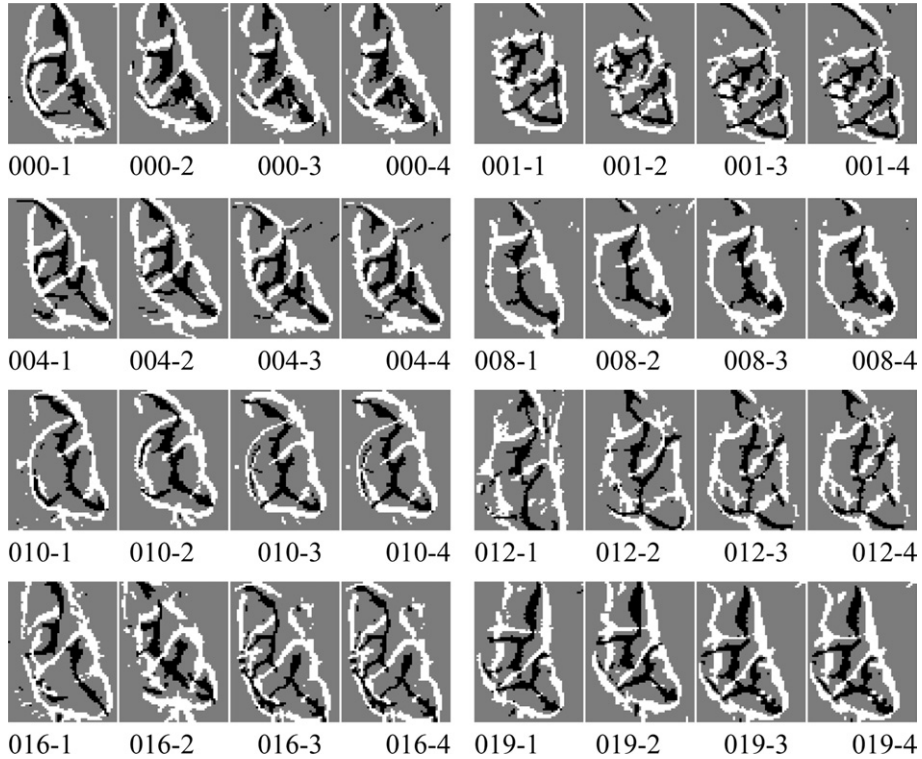


Fig. 11. A selection of thresholded convergence maps: four for each of eight subjects.

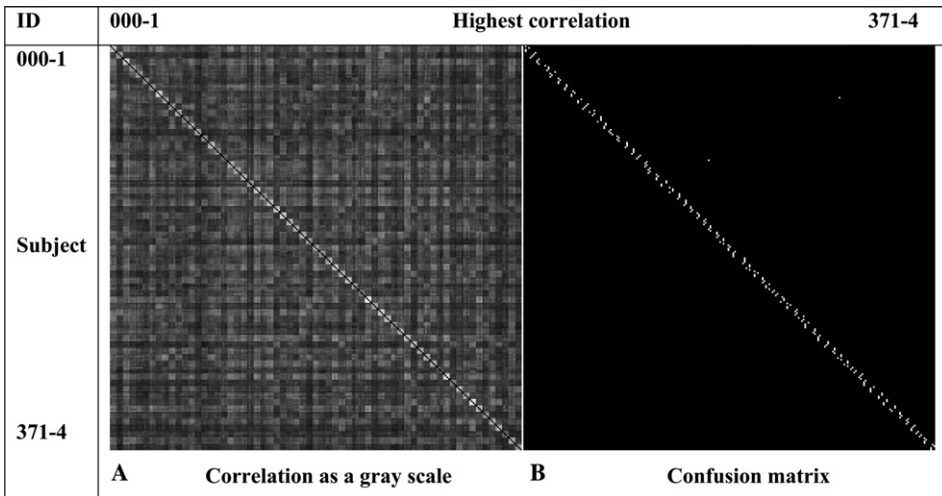


Fig. 12. Matrix of 252×252 cross-correlations.

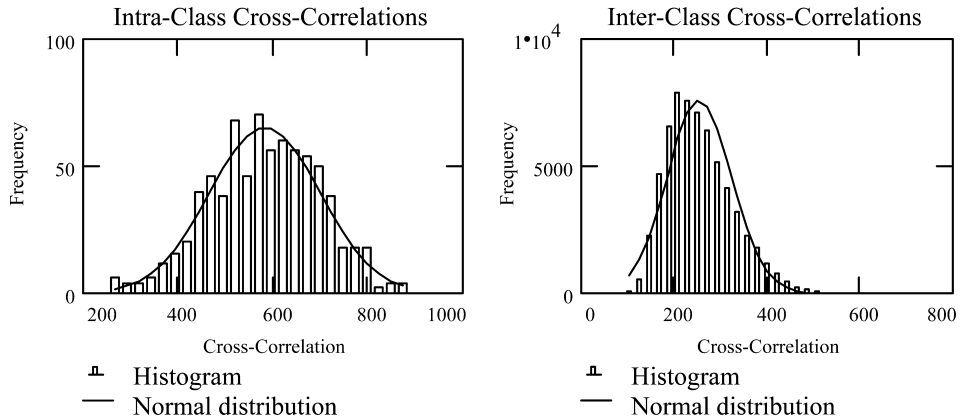


Fig. 13. Normal distribution fitted to frequency distribution of cross-correlations.

Since the frequency distributions are shown to be normally distributed, we can normalize them and form probability density functions as shown in Fig. 14.

The area of intersection formed between the two distributions, compared with the areas of the distributions themselves, gives an indication of the probability of misclassification.

The mean and standard deviation of the intra-class and inter-class correlations give a good indication of the nature of the results. Table 1 shows these statistics.

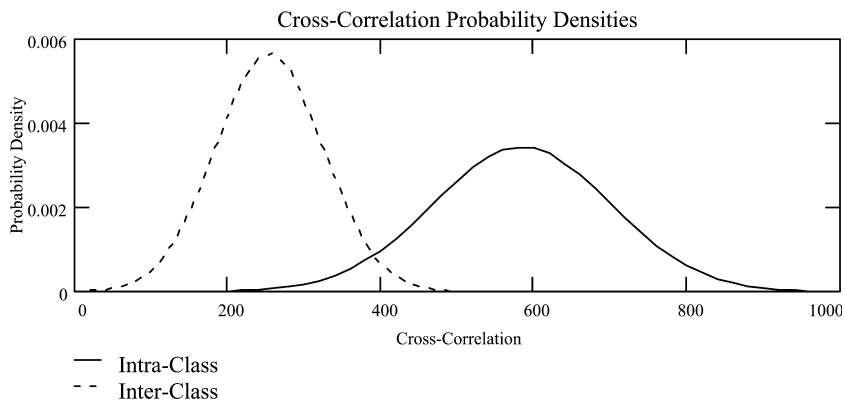


Fig. 14. Probability density functions of intra-class and inter-class correlations.

Table 1
Statistics for intra-class and inter-class convergence cross-correlations

Intra-class		Inter-class	
Mean	SD	Mean	SD
584.4	116.3	254.5	70.3

Note that the self-correlations that occur along the main diagonal of the cross-correlation matrix have been excluded from the calculations.

Using these results we may compute the decidability index after Daugman [30]. This index d' measures how well separated the distributions are, since recognition errors are caused by their overlap. The measure aims to give the highest scores to distributions with the widest separation between means, and smallest standard deviations. If the two means are μ_1 and μ_2 and the two standard deviations are σ_1 and σ_2 then d' is defined as

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{(\sigma_1^2 + \sigma_2^2)/2}}. \quad (13)$$

Substituting the means and standard deviations from Table 1, we obtain $d' = 3.43$.

4.4. Comparison with PCA

As promised earlier we now do a head-to-head comparison with the popular Principal Components Analysis technique [32]. The comparison is not quite head-to-head since FFE is symmetric in the sense that none of the samples are used for training, whereas we use 25% of the samples in forming the PCA covariance matrix. The tests are therefore less demanding for PCA since each subject is only compared with 188 others whereas for FFE they are compared against 251 other subjects.

The first member of each of the 63 classes is used to generate 63 eigenvectors which then form the basis vectors of the eigenspace. Fig. 15 shows the first four such eigenvectors. Each of the set of 252 images is projected onto the eigenspace, to obtain the corresponding projections. Fig. 16 shows the 4 projections for the first subject. Each projected image is made up of a linear combination of the basis vectors weighted by the projection components. Number 000-1 is naturally the most faithful reproduction since it has contributed to the subspace formation, but careful examination reveals minute differences when compared with Fig. 10A. As before, the projection vectors were exhaustively compared with each other, but this time it is the smallest Euclidian distance, rather than the highest cross-correlation which determines the best match. Not surprisingly, the results were poor because the ear images



Fig. 15. First 4 eigenvectors for 141×101 pixel images.

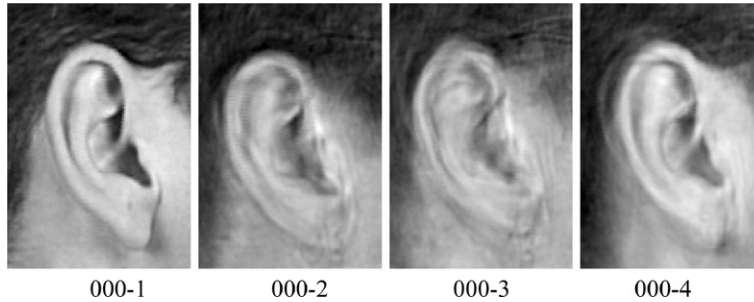


Fig. 16. First 4 projections for 141×101 pixel images.

are deliberately poorly registered. Out of the set of 189 images only 118, or 62.4%, gave correct matches and the decidability index was only 1.945.

To demonstrate that the bad results were due to poor extraction and registration and thereby highlight the inherent extraction advantage of the force field technique, the PCA experiment was run again but this time with the ears accurately extracted to the average data set ear size of 111×73 as shown in Fig. 17. This led to a dramatic improvement in the recognition rate to 98.4% and an index of 3.77 thus confirming the inherent extraction advantage of the force field approach. Notice that the index is slightly higher than the value of 3.43 obtained for the 141×101 images but this could be attributed to the reduction in data set size from 252 to 189 also to the fact that the images have been more fully extracted for PCA.

Fig. 18 shows the first 4 projections for the 111×73 pixel images where we notice a marked improvement in the quality of the projections compared with those in Fig. 16 due to the more accurate extraction and registration. The corresponding eigenvectors are shown in Fig. 19.

Finally, some experiments were conducted to assess the effect of brightness addition and the results compared with those for PCA under the same conditions. Each odd image has been subjected to the brightness change and every even image has been left unchanged, so that for each subject there are two bright images and two unchanged. These results are shown in tabular form in Table 2. where we see that the recognition rate for FFE falls by only 2% for a brightness addition of 3 standard

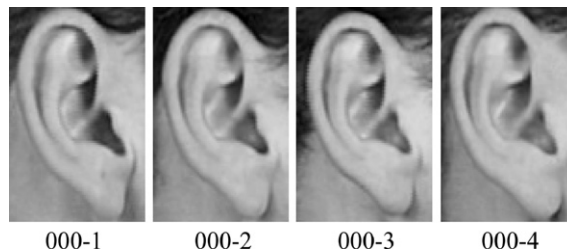


Fig. 17. Ear set cropped to average size of 111×73 .

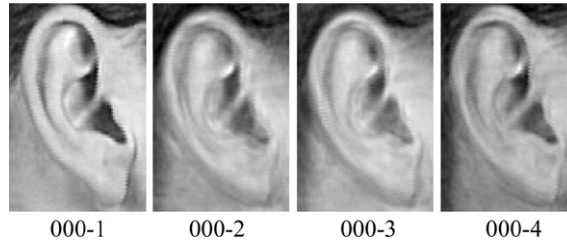
Fig. 18. First 4 projections for the 111×73 pixel images.Fig. 19. First 4 eigenvectors for 111×73 pixel images.

Table 2
Comparison of force field (FFE) and PCA recognition results

Image type	Method	Passes	Percentage (%)	Brightness addition	Decidability
141 × 101 with deliberately poor extraction and registration	FFE	250/252	99.2	None	3.432
	FFE	250/252	99.2	1 SD	3.384
	FFE	247/252	98.0	2 SD	3.137
	FFE	245/252	97.2	3 SD	2.846
	PCA	118/189	62.4	None	1.945
111 × 73 with accurate extraction and registration	PCA	186/189	98.4	None	3.774
	PCA	130/189	68.8	1 SD	1.694
	PCA	120/189	63.6	2 SD	0.878
	PCA	118/189	62.4	3 SD	0.476

deviations, compared to a fall of about 36% for PCA, thus confirming the robustness of FFE in the presence of variable lighting.

5. Conclusions

In conclusion we may say that in the context of ear biometrics we have developed a new linear transform that transforms an ear image, with very powerful smoothing and without loss of information, into a smooth dome shaped surface whose special shape facilitates a new form of feature extraction that extracts the essential ear

signature without the need for explicit ear extraction; and in the process we have verified the recognition potential of the human ear for biometrics. We have introduced the convergence operator and shown that it is a valuable alternative form of our earlier field line feature extraction. We have validated the technique by experiment and in the process we have contributed to the mounting evidence that ears are indeed viable as a biometric. In our future work we hope to refine our new technique and develop its full potential. We also hope to promote the case of ears as a new and promising biometric capable of competing in a rapidly developing field.

References

- [1] A. Bertillon, *La photographie judiciaire*, avec un appendice sur la classification et l'identification anthropométriques, Gauthier-Villars, Paris, 1890.
- [2] A. Iannarelli, *Ear Identification*, Paramount Publishing Company, Freemont, California, 1989.
- [3] STATE v. David Wayne KUNZE, Court of Appeals of Washington, Division 2. 97 Wash. App. 832, 988 P.2d 977, 1999.
- [4] Mark Dallagher Released, News item: The Chambers of William Clegg QC. Available from: <<http://www.2bedfordrow.co.uk/NewsDetail.asp?NewsID=17>>, 28 January 2004.
- [5] M. Burge, W. Burger, Ear biometrics, in: A. Jain, R. Bolle, S. Pankanti (Eds.), *Biometrics Personal Identification in Networked Society*, Kluwer Academic Publishers, 1999, pp. 273–286.
- [6] M. Burge, W. Burger, Ear biometrics in computer vision, in: *Proc. ICPR 2000*, 2002, pp. 822–826.
- [7] B. Victor, K. Bowyer, S. Sarkar, An evaluation of face and ear biometrics, in: *Proc. ICPR 2002*, 2002, pp. 429–432.
- [8] K. Chang, K.W. Bowyer, S. Sarkar, B. Victor, Comparison and combination of ear and face images in appearance-based biometrics, *IEEE Trans. PAMI* 25 (9) (2003) 1160–1165.
- [9] B. Moreno, A. Sanchez, J.F. Velez, On the use of outer ear images for personal identification in security applications, in: *Proc. IEEE 33rd Annual International Carnahan Conference on Security Technology*, 5–7 Oct. 1999.
- [10] D.J. Hurley, M.S. Nixon, J.N. Carter, Force field energy functionals for image feature extraction, in: *Proc. 10th British Machine Vision Conference BMVC99*, 1999, pp. 604–613.
- [11] D.J. Hurley, M.S. Nixon, J.N. Carter, Force field energy functionals for image feature extraction, *Image Vis. Comput.* 20 (2002) 311–317.
- [12] D.J. Hurley, Force field feature extraction for ear biometrics, PhD Thesis, University of Southampton, 2001.
- [13] D.J. Hurley, M.S. Nixon, J.N. Carter, A new force field transform for ear and face recognition, in: *Proc. IEEE Inter. Conf. on Image Processing ICIP2000*, 2000, pp. 25–28.
- [14] B. Luo, A.D. Cross, E.R. Hancock, Corner detection via topographic analysis of vector potential, *Pattern Recognit. Lett.* 20 (6) (1999) 635–650.
- [15] N. Ahuja, A transform for multiscale image segmentation by integrated edge and region detection, *IEEE Trans. PAMI* 18 (12) (1996) 1211–1235.
- [16] N. Ahuja, J.H. Chuang, Shape representation using a generalised potential field model, *IEEE Trans. PAMI* 19 (2) (1997) 169–176.
- [17] C. Xu, J.L. Prince, Gradient vector flow: a new external force for snakes, in: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 1997, pp. 66–71.
- [18] H.M. Yahia, J.P. Berroir, G. Mazars, Fast and robust level-set segmentation of deformable structures, in: *Inter. Conf. on Acoustics, Speech and Signal Processing, (ICASSP98)*, 1998.
- [19] K. Messer, J. Matas, J. Kittler, J. Luettin, G. Maitre, XM2VTSDB: the extended M2VTS Database, *Proc. AVBPA'99 Washington DC*, 1999.
- [20] B. Kimia, A. Tannenbaum, S.W. Zucker, Shapes, shocks, and deformations, I, *Int. J. Comput. Vis.* 15 (1995) 189–224.

- [21] K. Siddiqi, S. Bouix, A.R. Tannenbaum, S.W. Zucker, Hamilton-Jacobi skeletons, *Inter. J. Comput. Vis.* 48 (3) (2002) 215–231.
- [22] D. Halliday, R. Resnick, *Physics Part I*, third ed., Wiley, New York, 1977.
- [23] D. Halliday, R. Resnick, *Physics Part II*, Wiley, New York, 1962.
- [24] I.S. Grant, W.R. Phillips, *Electromagnetism*, Wiley, New York, 1990.
- [25] M.N.O. Sadiku, *Elements of Electromagnetics*, second ed., Saunders College Publishing, 1989.
- [26] G. Strang, *Linear algebra and its applications*, third ed., Saunders HBJ, 1988.
- [27] B. O’Neill, *Elementary Differential Geometry*, second ed., Physics [22,23], Academic Press, London and San Diego, CA, 1997.
- [28] P.J. Van Oterloo, *A Contour-oriented Approach to Shape Analysis*, Prentice-Hall, 1991.
- [29] S. Loncaric, A survey of shape analysis techniques, *Pattern Recognit.* 31 (8) (1998) 983–1001.
- [30] J. Daugman, *Biometric decision landscapes*, Technical Report TR482, University of Cambridge Computer Laboratory, 1999.
- [31] D. Marr, E. Hildreth, Theory of edge detection, *Proc. R. Sol. London B* 207 (1980) 187–217.
- [32] M. Turk, A. Pentland, Eigenfaces for recognition, *J. Cognit. Neurosci.* (1991) 71–86.