Force field energy functionals for image feature extraction

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

Department of Electronics and Computer Science, University of Southampton, Southampton S017 1BJ, UK

Abstract

The overall objective in defining feature space is to reduce the dimensionality of pattern space yet maintaining discriminatory power for classification and invariant description. To meet this objective, in the context of ear biometrics, a novel force field transformation has been developed in which the image is treated as an array of Gaussian attractors that act as the source of a force field. The directional properties of the force field are exploited to automatically locate the extrema of a small number of potential energy wells and associated potential channels. These form the basis of the ear description. This has been applied to a small database of ears and initial results show that the new approach has suitable performance attributes and shows promising results in automatic ear recognition. © 2002 Published by Elsevier Science B.V.

Keywords: Biometrics; Ear recognition; Force field transform; Potential energy transform; Potential channel

1. Introduction

In the context of machine vision, ear biometrics refers to the automatic measurement of distinctive ear features with a view to identify or confirm the identity of the owner. In many cases biometric studies are paralleled by forensic ones and ears have been part of forensic science for many years: a burglar was recently convicted of murder in the UK on the basis of ear prints found at the scene of the crime [1]; and an ear classification method has been developed for use in forensic science [2]. However, as a biometric it has received scant attention compared with the more popular techniques of automatic face, eye, or fingerprint recognition, though one automated system for ear identification has been developed recently [3]. There is also anecdotal evidence of military interest in ear recognition, though no works have appeared in the public domain.

An ear recognition system could be used just like other biometric systems, say for access control. A database or register would be prepared by processing images of the ears of authorized personnel to extract a set of characteristic features for each image. Personnel wishing to enter would have their ears scanned at the entrance and the image would be processed and compared for a match against the database. The stored feature vectors would have to be sufficiently distinct so as to be able to distinguish one ear from all the others and sufficiently robust so that the same vector would be produced every time the ear is scanned. These are conflicting requirements and present a challenge to the system designer.

There are a number of techniques with potential to find and describe a human ear by computer vision. Essentially, we need to find an ear and describe it for recognition. Clearly, there are application constraints, such as occlusion by hair, but here we are concerned with basic technique. Ear extraction could use an active contour [4] but with initialisation problems which can be relieved by a dual active contour [5] (though this still requires establishment of inner and outer contours). The only published approach to ear biometrics [3] modelled ears as an adjacency graph built from a Voronoi diagram of its curve segments. Then, error correcting graph matching was applied to remove false curves, but recognition performance on a database was not included. Techniques derived from fingerprint analysis or texture classification could also be used to describe the folds and ridges in a human ear. To give appropriate description capability, a novel two-stage approach has been developed to provide ear extraction and description concurrently in a reliable and robust manner. The two stages are: *image to force field transformation*; and *potential well and channel extraction* [6].

The entire image is converted into a force field by assuming, purely as a mathematical convenience, that each pixel exerts an isotropic force on all the other pixels that is proportional to pixel intensity and inversely proportional to the square of the distance, i.e. the inverse square law.

---

* Corresponding author.

E-mail addresses: djh@hurley25.fsnet.co.uk (D.J. Hurley), msn@ecs. soton.ac.uk (M.S. Nixon), jnc@ecs.soton.ac.uk (J.N. Carter).

There is a potential energy surface associated with this force field, which in the case of an ear can be likened to a small mountain with a few peaks joined by ridges. We call these peaks potential energy wells and extending the well analogy we refer to the ridges that lead to the peaks as potential energy channels.

The structure of the force field, as described by the field lines, shows remarkable invariance in that if the initialising position is translated or the image is scaled then the same description will result. We also find that the process is very tolerant of noise, due to its inherent averaging. We shall describe the force field approach in Section 2, demonstrating its originality and its invariance properties. In Section 3 we will show how these performance attributes are satisfied in images of ears, together with analysis of recognition capability on a small database of ears, prior to further work and conclusions.

2. Force field approach

Mathematical modelling techniques used in physics have recently attracted the attention of researchers in computer vision; for example Ref. [9] describes the use of vector potential to extract corners by treating the Canny edge map of the image as a current density. Another recent approach [10] has used a potential field model in a medial axis transform. The concepts underpinning the force field transformation and the mathematics used to describe it can be found in various introductory works on physics [7] and electromagnetics [8].

The image is transformed by assuming, purely as a mathematical convenience, that it consists of an array of N attractive particles that act as the source of a force field. Each pixel is considered to generate a spherically symmetric force field so that the force $\mathbf{F}_i(\mathbf{r}_j)$ exerted on a pixel of unit intensity at the pixel location with position vector $\mathbf{r}_j$ by any other pixel with position vector $\mathbf{r}_i$ and pixel intensity $P(\mathbf{r}_i)$ is given by

$$\mathbf{F}_i(\mathbf{r}_j) = P(\mathbf{r}_i)\frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} \tag{1}$$

This equation shows that the magnitude of the force is governed by the inverse square law and that there is also direction: the force is a vector, not a scalar. The units of pixel intensity, force, and distance are arbitrary, as are the co-ordinates of the origin of the vector field. The total force $\mathbf{F}(\mathbf{r}_j)$ exerted on a pixel of unit intensity at the pixel location with position vector $\mathbf{r}_j$ is the vector sum of all the forces due to the other pixels in the image and is given by

$$\mathbf{F}(\mathbf{r}_j) = \sum_{i \in 0, N-1 | i \neq j} \mathbf{F}_i(\mathbf{r}_j) = \sum_{i \in 0, N-1 | i \neq j} \left(P(\mathbf{r}_i)\frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} \right) \tag{2}$$

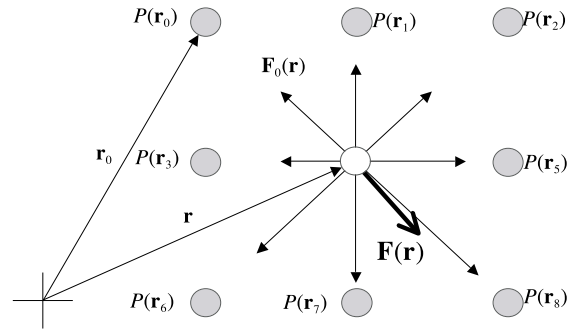In order to calculate the force field for the entire image, this

Fig. 1. Force field calculation at a single pixel position.

equation should be applied at every pixel position in the image. This calculation is illustrated graphically for the total force acting at a typical pixel position in Fig. 1.

There is a spherically symmetric scalar potential energy field associated with the force field generated by each pixel where $E_i(\mathbf{r}_j)$ is the potential energy imparted to a pixel of unit intensity at the pixel location with position vector $\mathbf{r}_j$ by the energy field of any other pixel with position vector $\mathbf{r}_i$ and pixel intensity $P(\mathbf{r}_i)$, and is given by

$$E_i(\mathbf{r}_j) = \frac{P(\mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{3}$$

The defining equation is simpler than the force field equation but the concept is less intuitive. If an exploratory unit test pixel is moved around in the force field generated by a given pixel, energy will be exchanged if the net effect is to change the distance of the test pixel from the given pixel. Thus, in the image plane, the field consists of concentric rings of equal potential energy known as equipotentials. If the test pixel moves to a different location on the same equipotential ring, no energy is exchanged. If it moves to a different equipotential, an amount of energy will be exchanged equal to the difference in potential energy between the two rings. The potential energy function of a single isolated pixel is shown in Fig. 2.

Now to find the total potential energy at a particular pixel location in the image, the scalar sum is taken of the values of the overlapping potential energy functions of all the image pixels at that precise location and is given by

$$E(\mathbf{r}_j) = \sum_{i \in 0, N-1 | i \neq j} E_i(\mathbf{r}_j) = \sum_{i \in 0, N-1 | i \neq j} \frac{P(\mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{4}$$

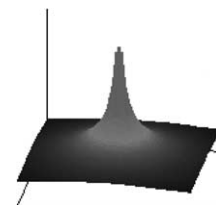This summation is then carried out at each pixel location to

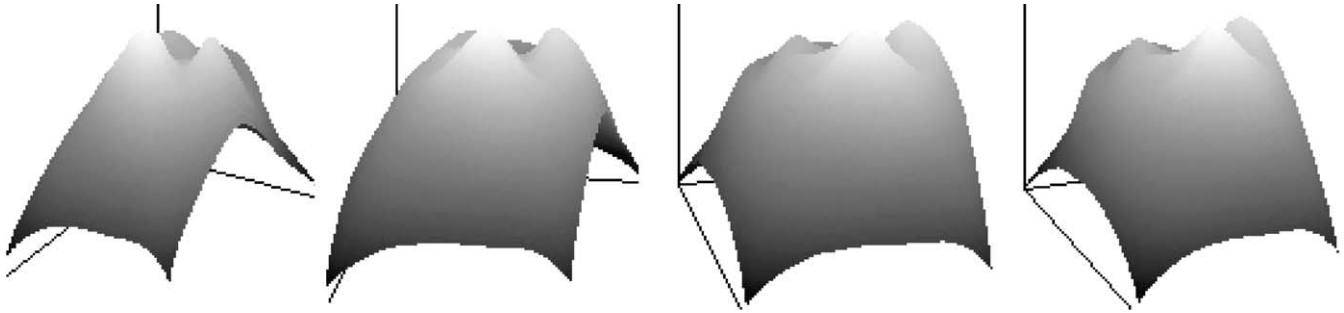Fig. 2. Isolated pixel potential energy function.

Fig. 3. Potential energy surface for a human ear.

generate a potential energy surface, which is a smoothly varying surface due to the fact that the underlying inverted vortices have smooth surfaces. The potential energy surface for an ear is shown in Fig. 3.

In application, an array of unit value exploratory mobile test pixels is arranged in a closed loop formation surrounding the target ear. Each test pixel is then allowed to follow the pull of the force field so that its trajectory forms a field line and it will continue moving until it reaches an extremum in the potential energy surface. The gradient is zero at an extremum so no force is exerted and no further movement is possible. Since the force field at a point is unique, all field lines that arrive at a given point will follow the same path from that point onwards thus forming channels. This process is illustrated in Fig. 4 where an elliptic array of 50 test pixels is placed in the force field and iterated to produce field lines as shown. The most striking example of the channel formation process is seen at the top of the ear where 14 field lines combine to form a channel which flows rightwards following the contour of the ear-rim. The locations of the minima (or wells) are extracted by simply noting the coordinates of the clusters of the test pixels eventually form. These locations are shown on the right of Fig. 4, superimposed on the force field magnitude.

For the approach to be of practical use, we require to show initialisation and scale invariance. If we can show scale invariance then initialisation invariance follows naturally since test pixels will trace out the same field lines albeit scaled-down versions. Accordingly, we will show that should the image be scaled by a given scale factor then the force and energy fields scale by the same factor. By the nature of the underlying equations, we can analyse one property and the results generalise naturally to the other.

Should the image be scaled by a factor s then we have that

$$E(s\mathbf{r}_j) = \sum_{i \in 0,N-1 | i \neq j} E_i(s\mathbf{r}_j) = \sum_{i \in 0,N-1 | i \neq j} \frac{P(s\mathbf{r}_i)}{|s\mathbf{r}_i - s\mathbf{r}_j|}$$

$$= \frac{1}{s} \sum_{i \in 0,N-1 | i \neq j} \frac{P(s\mathbf{r}_i)}{\mathbf{r}_i - \mathbf{r}_j} \tag{5}$$

We see that scaling the image distribution not only results in a scaled force field distribution but, as expected, larger distances imply weaker forces as the denominator is proportional to scale. As such, the resulting force and potential fields are invariant to scale.

Next we seek to demonstrate tolerance to noise. Given that the image is corrupted by additive zero-mean Gaussian noise ν we have that

$$E(\mathbf{r}_j) = \sum_{i \in 0,N-1 | i \neq j} E_i(\mathbf{r}_j) = \sum_{i \in 0,N-1 | i \neq j} \frac{P(\mathbf{r}_i) + \nu_i}{|\mathbf{r}_i - \mathbf{r}_j|}$$

$$= \sum_{i \in 0,N-1 | i \neq j} \frac{P(\mathbf{r}_j)}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i \in 0,N-1 | i \neq j} \frac{\nu_i}{|\mathbf{r}_i - \mathbf{r}_j|} \tag{6}$$

Given that the noise ν is zero-mean then the energy field is
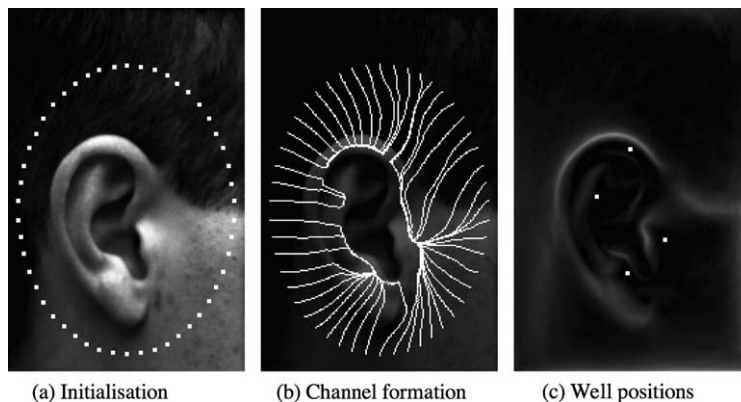
(a) Initialisation   (b) Channel formation   (c) Well positions

Fig. 4. Extraction of potential wells and channels.

unaffected by noise, so long as the noise is mean-zero over the image. Naturally, noise with high variance will violate this assumption so it appears possible to add extreme noise and the technique will fail, as expected. This is the averaging inherent in the transformation, as referred to earlier. Should the image intensity be scaled by a factor k then we have that

$$E(\mathbf{r}_j) = \sum_{i \in 0, N-1 | i \neq j} E_i(\mathbf{r}_j) = \sum_{i \in 0, N-1 | i \neq j} \frac{kP(\mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_j|}$$

$$= k \sum_{i \in 0, N-1 | i \neq j} \frac{P(\mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_j|} \qquad (7)$$

We see that scaling the image intensity does not alter the resulting force field distribution but, as expected, only scales the energy intensity by the same factor. As such, the resulting force and potential fields are invariant under multiplicative changes in illumination.

Similarly, change in overall brightness by addition gives

$$E(\mathbf{r}_j) = \sum_{i \in 0, N-1 | i \neq j} E_i(\mathbf{r}_j) = \sum_{i \in 0, N-1 | i \neq j} \frac{P(\mathbf{r}_i) + k}{|\mathbf{r}_i - \mathbf{r}_j|}$$

$$= \sum_{i \in 0, N-1 | i \neq j} \frac{P(\mathbf{r}_i)}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i \in 0, N-1 | i \neq j} \frac{k}{|\mathbf{r}_i - \mathbf{r}_j|} \qquad (8)$$

This suggests that addition of k brightness levels to each pixel results in an increment to the steepness of the surface shown in Fig. 3. This will have the effect of a slight non-linear scale stretching of the feature distribution, but otherwise will preserve its form. Naturally, the force field will be affected by localized changes in illumination.

3. Invertible linear transform

We show that the force field transform is a linear transformation by developing its matrix representation. The form of the matrix is illustrated for a trivial 2×2 pixel image. It is easily verified that this represents the application of Eq. (2) at each of the four pixel locations. This equation multiplies a column vector of pixel intensities, P_i, by a matrix of inverse square displacement vectors $\mathbf{d}_{ij}$ to give a column vector of

forces, $\mathbf{F}_i$. We have

$$\begin{pmatrix} 0 & \mathbf{d}_{01} & \mathbf{d}_{02} & \mathbf{d}_{03} \\ \mathbf{d}_{10} & 0 & \mathbf{d}_{12} & \mathbf{d}_{13} \\ \mathbf{d}_{20} & \mathbf{d}_{21} & 0 & \mathbf{d}_{23} \\ \mathbf{d}_{30} & \mathbf{d}_{31} & \mathbf{d}_{32} & 0 \end{pmatrix} \begin{pmatrix} P_0 \\ P_1 \\ P_2 \\ P_3 \end{pmatrix} = \begin{pmatrix} \mathbf{F}_0 \\ \mathbf{F}_1 \\ \mathbf{F}_2 \\ \mathbf{F}_3 \end{pmatrix} \qquad (9)$$

where

$$\mathbf{d}_{ij} = \frac{\mathbf{r}_j - \mathbf{r}_i}{|\mathbf{r}_j - \mathbf{r}_i|^3}$$

This is a skew-symmetric matrix: the leading diagonal of zeros reflects the fact that no pixel attracts itself and the skew symmetry is accounted for by the fact that we are dealing with a fully connected network but with a pair of directed edges connecting every pair of nodes.

There is a corresponding representation for the potential energy transformation since the vector force field and scalar potential energy fields are related by the fact that the force at a given point is equal to the additive inverse of the gradient of the potential energy surface at that point,

$$\mathbf{F}(\mathbf{r}) = -\text{grad}(E(\mathbf{r})) = -\nabla E(\mathbf{r}) \qquad (10)$$

Since the representation matrices are square it is of theoretical interest whether they are invertible or not. If they are invertible then the original image can be recovered for example from the potential energy surface. This implies that all the information in the original image is conserved by the transformation, which is an important result. In practice the representation matrices for images of even modest size are very large, for example a 10×10 image has a matrix with 10,000 elements. However, we have tested the potential energy representation matrices for all square images up to 32×32 pixels and all non-square images up to 7×8 pixels and have found them to be invertible. These results suggest that the potential energy transform is indeed invertible for most image sizes and aspect ratios. Even if there are some particular combinations of aspect ratio and size that yield singular matrices, this should not detract from the overall conclusion that all of the information is conserved by the transformation.
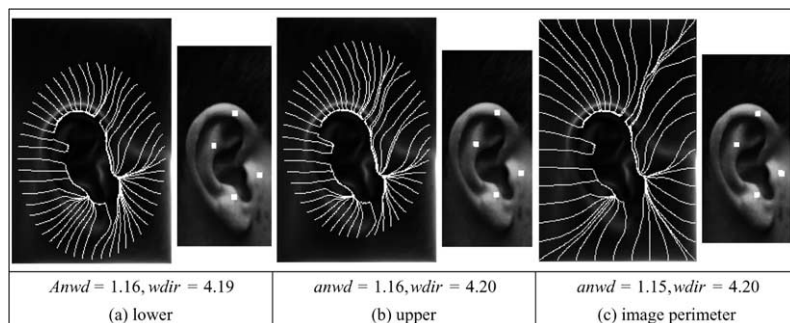

$Anwd = 1.16, wdir = 4.19$	$anwd = 1.16, wdir = 4.20$	$anwd = 1.15, wdir = 4.20$
(a) lower	(b) upper	(c) image perimeter

Fig. 5. Initialisation invariance.

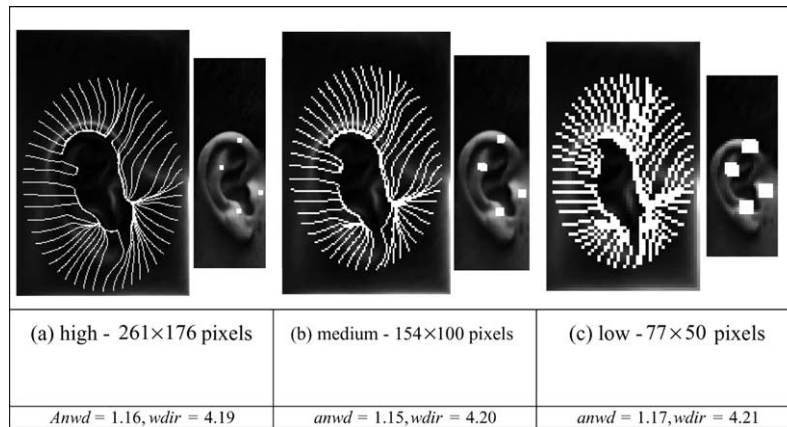| (a) high - 261×176 pixels	(b) medium - 154×100 pixels	(c) low - 77×50 pixels
Anwd = 1.16, *wdir* = 4.19	*anwd* = 1.15, *wdir* = 4.20	*anwd* = 1.17, *wdir* = 4.21

Fig. 6. Scale invariance.

4. Ear analysis by force field transformation

Having demonstrated how field lines can be employed to extract potential wells and channels, we now demonstrate that the new technique is robust. We do this by demonstrating that the technique can indeed exhibit initialisation, scale, and translation invariance, as predicted. We also show that the descriptions derived from positions of the wells are unique for different ears, and show requisite performance attributes described earlier. To assess difference between the descriptions, we shall use a measure of the average normalised distance of the well positions *anwd*, together with the accumulated direction to the position of each well-point from a chosen reference point, *wdir*. For W wells at points w_i, these measures are:

$$\text{anwd} = \frac{\sum_{i=1}^{W} |\mathbf{w}_i|}{\max(|\mathbf{w}_i|)W} \tag{11}$$

and

$$\text{wdir} = \sum_{i=1}^{W} (\langle \mathbf{w}_i) \tag{12}$$

Fig. 5 demonstrates initialisation invariance, confirming the result of Eq. (5). The measures are shown for each different starting point, and show that very similar measures are achieved, reflecting visual analysis of these results. The centre of the ellipse has been displaced downwards in the left hand image and upwards in the middle image. The force field structure is essentially preserved across the three images and the location of the potential wells is the same in each case. The rightmost image shows an initialisation along the borders of the image at intervals of 20 pixels. Naturally, it is chance whether the ellipse starting points coincide with one of the field lines. This is illustrated in the central image where a left hand well is formed from three field lines, as opposed to two for the other initialisations. Clearly the result could vary with the number of ellipse points, especially if the number was very small where wells naturally would be omitted as no field lines would converge to them. This suggests simply that the number of initialisation points should be sufficiently large and widely distributed so as to ensure that all wells will be extracted.

Fig. 6 demonstrates scale invariance, confirming the result of Eq. (6) (but note that quantisation effects were not included in the earlier analysis). The descriptions of the wells produce very similar results. The structure of the force field is preserved when an image is captured and transformed at lower resolution. The images decrease in resolution from left to right, being 261 × 176, 154 × 100
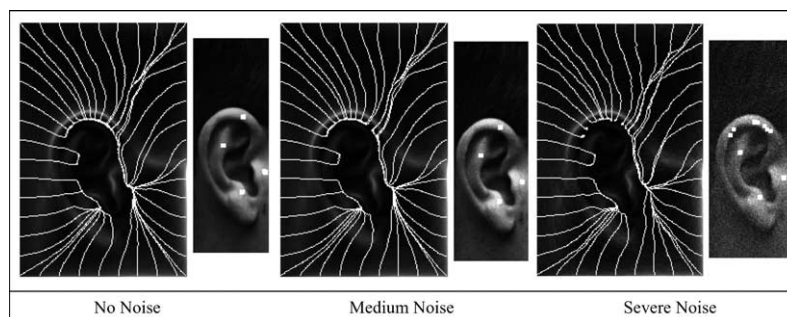

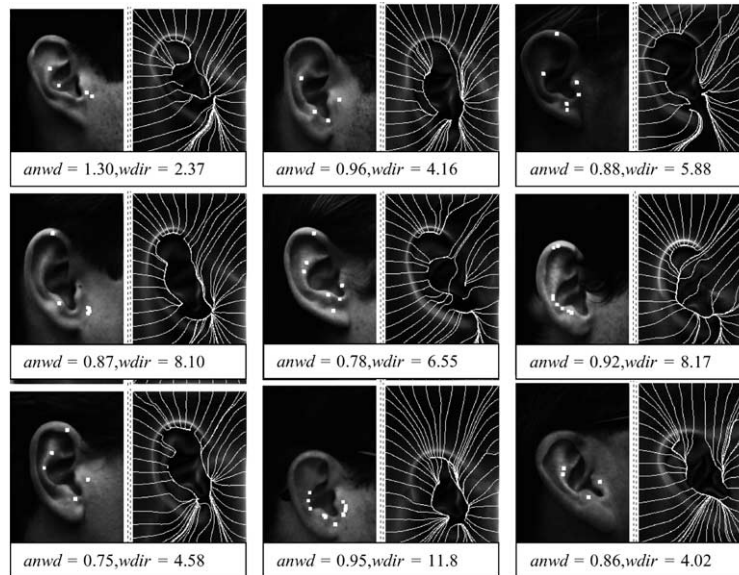Fig. 7. Illustration of noise tolerance.

Fig. 8. Uniqueness of well and channel descriptions.

5. Conclusions and further work

and 77×50 pixels, respectively. This is important since it implies that scale space techniques can be employed so that a low-resolution image could be used to locate a target's position and a higher resolution version could then be used to refine feature information.

Noise tolerance is demonstrated in Fig. 7 where it is seen that the force field structure is essentially preserved in the presence of Gaussian noise. Notice that even though the channels begin to break up into individual wells in the presence of severe noise, that the channel outline is still clearly discernible. By this, an analysis of the position of the wells is inappropriate here, but a description of the channels is likely to produce similar results, for the same ear, when the noise increases.

Fig. 8 demonstrates that nine different ears produce quite different feature vectors and that potential channels and well locations are unique to each image. Further, the measures are quite different to those for Figs. 5 and 6 (which are those for a different ear), which show that this small database can indeed be discriminated by these measures. Clearly, a richer selection of measures will emphasise this effect. Note that in the lower right part of the ear, channel formation appears to behave with medial axis properties; the channels' course being influenced both by the corner of the notch and the rim to the left.

Future work, on a large database, will aim to confirm the potential for this technique in ear recognition aiming not only to use the positions of the wells, but also to investigate the potential of channel description for recognition purposes. Clearly, using ears as a biometric is at an exploratory stage and further work will seek to develop a database that includes variation in sex, age, ethnic origin, jewellery, hair and other factors.

We have developed a new feature extraction technique, targeted primarily at ear biometrics, with remarkable invariance to initialisation and scaling, and which demonstrates good noise tolerance and promising recognition capability. The beauty of this technique is that an explicit description of the ear topology is not necessary and extracting the ear biometric is simplicity by itself—merely follows the force field lines and observe eventual clustering of coordinates. Taking account of the channel shape and ultimately the underlying shape of the energy surface will increase the level of detail in the description to meet any demand.

Whilst the force field transform has been demonstrated in the context of ear biometrics, we feel that it is an important new development in its own right. We aim later to investigate its potential in other application domains. An important aspect of the transformation is the fact that it is similar to a natural process, namely the formation of electric fields in the vicinity of electric charge distributions, suggesting the prospect of a solid state device with direct image to force field conversion in real time. Naturally, as the present approach is computationally demanding we already aim to develop means to improve its computation.

We have shown that the force field transform can be applied in two dimensions to assist with ear recognition. However, since the formulation is couched in vector form, the concept could generalise to higher dimensional problems. Whilst the two dimensional case is easily depicted, the N-dimensional case can only be interpreted numerically. Full field solutions have been used to date to establish a reliable baseline against which approximate solutions can be compared and we aim to investigate its generalisation capability to higher dimensions.

References

[1] BBC News, Ear print catches murderer, BBC Online Network, 15 Dec 1998.

[2] A. Iannarelli, Ear Identification, Paramount Publishing Company, Freemont, CA, 1989.

[3] M. Burge, W. Burger, Ear identification, in: A.K. Jain, R. Bolle, S. Pankanti (Eds.), Biometrics: Personal Identification in Networked Society, Kluwer Academic, Dordrecht, 1998, pp. 273–286.

[4] M. Kass, A. Witkin, D. Terzopoulos, Snakes: active contour models, International Journal of Computer Vision 1 (1988) 321–331.

[5] S.R. Gunn, M.S. Nixon, A robust snake implementation; a dual active contour, IEEE Transactions PAMI 19 (1) (1997) 63–67.

[6] D.J. Hurley, M.S. Nixon, J.N. Carter, Force field energy functionals for image feature extraction, Proceedings of the Tenth British Machine Vision Conference BMVC99, 2 (1999) 604–613.

[7] D. Halliday, R. Resnick, Physics Part II 1962 Wiley International Edition, Wiley, New York.

[8] I.S. Grant, W.R. Phillips, Electromagnetism, Second ed., Wiley, New York, 1990.

[9] B. Luo, A.D. Cross, E.R. Hancock, Corner detection via topographic analysis of vector potential, Proceedings of the Ninth British Machine Vision Conference BMVC98, 2 (1998) 567–576.

[10] N. Ahuja, J.H. Chuang, Shape representation using a generalized potential field model, IEEE Transactions PAMI 19 (2) (1997) 169–176.