Copyright © 2006IEEE.

Reprinted from the Third International Symposium on 3D Data Processing,
Visualization and Transmission (3DPVT 2006).

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the products or services of The University of Edinburgh, School of Informatics. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

Qualitative Characterization of Deforming Surfaces

T.C. Lukins and R.B. Fisher
School of Informatics
University of Edinburgh
{s0126209,rbf}@inf.ed.ac.uk

Abstract

This paper extends the idea of classification schemes for static surface curvature into the temporal domain. We seek to identify regions in sequences of depth data that exhibit variations in shape change, and to characterise the nature of the deformation. From observing the change in principle curvatures we show how it is possible to decouple the type of change into one of fifteen classes, and also reveal the extent of alteration. Results are presented for synthetic and real data sequences, with additional alignment performed to accommodate global motion. This technique shows promise in analysing data from video-rate range sensors, with potential applications in biometric and psychological analysis of the face and other deformable objects.

1. Introduction

The study of shape is most often referred to as *morphometrics* [4]. This has traditionally been the domain of biological taxonomy, broadly divided into biometric and geometric based approaches. The biometric approach seeks to identify specific landmarks and compare the measurements between them (e.g. by using the Procrustes transform). In contrast, the more modern geometric approach tries to preserve a more complete description of the shape (e.g. by using thin-plate spline and Finite Element Model based representations). The aim in both cases is to provide **quantitative** assessment of the variability of form in response to other experimental conditions. This must take account of the fact that the description of solid form is itself fundamentally not only a property of the object, but is also based upon the operation (method) or representation (model) used [8]. A common problem is then in recognising the same conceptual terms in order to express equivalence, given that there are many different examples of a particular shape. Such formalism is particularly relevant when we seek to not only *capture*, but to also *describe* the dynamic nature of the transitions between forms.

In this paper we introduce a vocabulary and a computational technique that makes explicit the deformations of a surface over time. We show that this can be used to differentiate between the *type of deformation* and the *extent of change*. In so doing, we maintain that this produces a compact representation of dynamics for **qualitative** characterisation of deforming surfaces. The calculations for this can be performed by applying techniques from differential geometry to $4D$ data sequences.

1.1. Capturing Deformation

A great deal of the work in capturing dynamic shape in Computer Vision is concerned with medical imaging. In particular, for describing the alteration in morphology for various organs of the body over time - especially the brain and heart. Such work is mostly based on volumetric data from MRI or CAT scans that must first be segmented and registered. In particular, the alignment of data must attempt to accommodate those regions which have actually deformed, as opposed to those area that have remained constant. In this respect, the local properties of the surface can be useful in identifying change - for example normals and curvature - and can be used to refine the fitting between scans [5]. The non-linear nature of natural tissue makes subsequent modelling particularly hard to achieve, most commonly employing Finite Element Models [14] and displacement fields [6] to track and capture the deformation.

Another popular domain that seeks to model deformation is the analysis of the human face. Research in this field is based primarily on data gained from $3D$ acquisition, again with the initial focus in fitting a representative model in order to accommodate registration. If the alignment is correct, then comparing two different scans can be made, where invariant properties of the surface can again be particularly useful in recognising the same person regardless of expression. For example, by using shape curvature to find features for an initial coarse alignment between two facial scans, before then refining the match with the Iterative Closest Point algorithm in order to identify the person [11].

A more holistic approach can compare two faces using the thin plate spline representation (common in geometric morphometrics) to estimate the actual non-rigid displacement vector field, and so delineate the deformation and isolate areas that have not changed [12]. In such approaches, the emphasis is again focused on tracking the correspondences between multiple scans of an object. Other work for ensuring accurate comparative analysis of scans relates back to more landmark oriented schemes, coupled to physical elasticity modelling as a natural extension to biometric morphometrics [20].

The modelling of dynamic deformations over shorter (millisecond) timescales can be further related to this problem of tracking the motion of corresponding points. In image based techniques it is common to use optic flow, with the possibility to decompose the resulting vector fields into component elements - such as gradient, divergence and curl - that reveal something of the type of motion [9]. However, the extension of these techniques into $3D$ (for example via scene and range-flow [16]) currently serve to only capture localised changes. In this respect, hybrid flow based techniques [13], as well as harmonic maps [17], can be employed to help constrain the newly emerging $4D$ capture technologies for space-time reconstruction [19]. This has lead to the possibility of video-rate capture of range data for new commercial and academic systems [18].

While these recent advances allow for capture of actual dynamic surfaces, there has not been much work to date that seeks to analyse, nor describe, the observed deformations. Indeed, we note in general that there is an absence of common terminology, or any complete geometric definition. This motivates to to focus on producing a simplified and more symbolic technique for presenting the dynamics of the data in a meaningful and accessible way.

1.2. Static Curvature Descriptions

The curvature of a surface is an intuitive idea to describe the nature of shape and deviation from flatness of an embedded manifold in $\mathbb{R}^3$. In more formal differential geometry terms these properties are defined by considering the local intersection of a tangent plane with the surface at a fixed point. Varying the orientation of this plane around a normal vector will reveal two orthogonal extrema - representing the two *principal curvature* values (κ_1 and κ_2) and the directions at which they occur as the *principal directions*.

Exploiting this viewpoint invariant basis for curvature has a long history of use in Computer Vision as a means of object recognition. The most widely known scheme was introduced by Besl in 1986 based on a *mean* ($H = \frac{1}{2}(\kappa_1 + \kappa_2)$) and *Gaussian* ($K = \kappa_1 * \kappa_2$) curvature based segmentation for 8 possible - and 1 impossible - shape classifications [2].

While this discretization to a subset of shapes is useful, such a scheme loses the means to describe the degree to which the surface is actually curved, and the progression in shape. In response to this Koenderink proposed an alternative polar based representation which directly decouples a continuous *shape* ($S = \frac{2}{\pi} arctan\left((\kappa_1 + \kappa_2)/(\kappa_1 - \kappa_2)\right)$) from an independent value of *curvedness* ($C = \sqrt{(\kappa_1^2 + \kappa_2^2)/2}$) [10] - as shown in figure 1. Both the $H + K$ and $S + C$ techniques continue to have many applications in analysing static range/depth data from scenes, and inspire us to consider how they might be extended to apply over time.

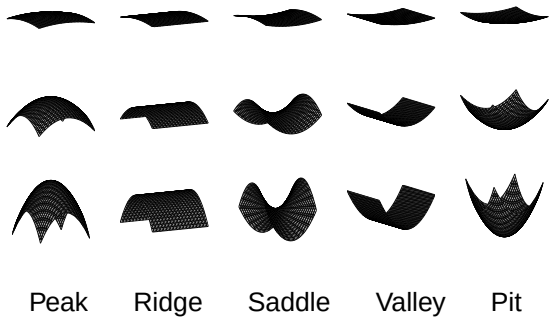


Figure 1. Shape and curvedness.

2. Formulation

We propose a natural extension to the classification of surface deformation by observing the variation of the principal curvature values over time. As illustrated in the previous section, various classification schemes are available for describing a static surface. In particular, the primary 6 classes of *prototypical shape* which are possible via comparison of the relative signs for the principal curvatures κ_1 and κ_2 are shown in Table 1.

	$\kappa_1 < 0$	$\kappa_1 = 0$	$\kappa_1 > 0$
$\kappa_2 < 0$	concave ellipsoid ("pit")	concave cylinder ("valley")	hyperboloid ("saddle")
$\kappa_2 = 0$	concave cylinder ("valley")	plane ("flat")	convex cylinder ("ridge")
$\kappa_2 > 0$	hyperboloid ("saddle")	convex cylinder ("ridge")	convex ellipsoid ("peak")

Table 1. Shape classes based on κ_1 and κ_2.

Initial Shape:		$\kappa_1 < 0$			$\kappa_1 = 0$			$\kappa_1 > 0$		
Dynamic Change:		$\Delta\kappa_1 < 0$	$\Delta\kappa_1 = 0$	$\Delta\kappa_1 > 0$	$\Delta\kappa_1 < 0$	$\Delta\kappa_1 = 0$	$\Delta\kappa_1 > 0$	$\Delta\kappa_1 < 0$	$\Delta\kappa_1 = 0$	$\Delta\kappa_1 > 0$
	$\Delta\kappa_2 < 0$	Subside	Squeeze	Squeeze	Collapse	Fold	Dimple	Crumple	Crumple	Warp
$\kappa_2 < 0$	$\Delta\kappa_2 = 0$	Squeeze	Constant	Squeeze	Collapse	Constant	Dimple	Crumple	Constant	Crease
	$\Delta\kappa_2 > 0$	Squeeze	Squeeze	Flatten	Collapse	Flatten	Dimple	Flatten	Crease	Crease
	$\Delta\kappa_2 < 0$	Collapse	Collapse	Collapse	Subside	Fold	Warp	Dent	Dent	Dent
$\kappa_2 = 0$	$\Delta\kappa_2 = 0$	Fold	Constant	Flatten	Fold	Constant	Bend	Flatten	Constant	Bend
	$\Delta\kappa_2 > 0$	Dimple	Dimple	Dimple	Warp	Bend	Protrude	Bulge	Bulge	Bulge
	$\Delta\kappa_2 < 0$	Crumple	Crumple	Flatten	Dent	Flatten	Bulge	Flatten	Stretch	Stretch
$\kappa_2 > 0$	$\Delta\kappa_2 = 0$	Crumple	Constant	Crease	Dent	Constant	Bulge	Stretch	Constant	Stretch
	$\Delta\kappa_2 > 0$	Warp	Crease	Crease	Dent	Bend	Bulge	Stretch	Stretch	Protrude

Table 2. The 15 types of deformation based on change in principal curvatures from initial shape.

For these classifications to work, it is often necessary to define the zero boundary regions with a threshold ($\kappa = 0 \iff -\theta_{shape} < \kappa < \theta_{shape}$). If one then considers the alteration of κ_1 and κ_2 over time, then the transitions that can occur from one prototype to another, and the dynamic relationships between the classes, can be visualised as the graph shown in Figure 2.

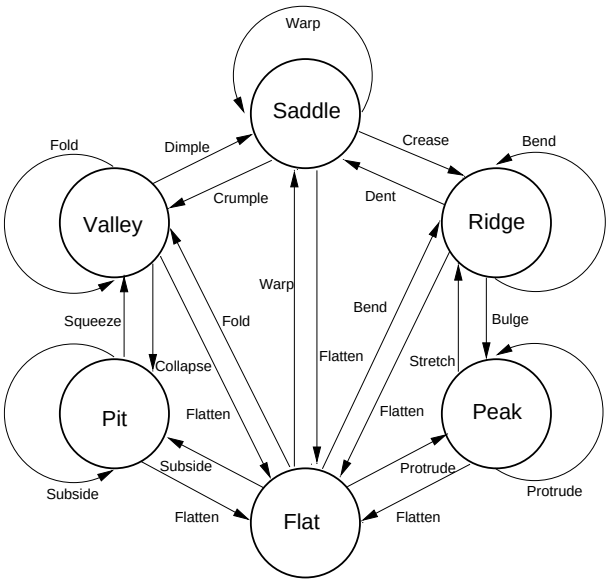


Figure 2. Transitions between shapes.

It is evident from Figure 2 that it is firstly impossible for some shape classes to deform into others without first transitioning through an intermediate form (e.g. for a peak to turn to a valley, it must first become flat, or else move from a ridge to a saddle). Secondly, it can be realised that there are only a limited number of types of transition:

- Those that typify, and exemplify further, the formation of a prototype (e.g. "**protrude**" a peak, "**subside**" a pit, "**fold**" a valley, "**bend**" a ridge, "**warp**" a saddle).
- Those that move the opposite way from the prototype towards flat (e.g. "**flatten**").
- Those bi-directional transitions that are also applicable only between neighbouring non-flat prototypes (e.g. to "**squeeze**" a pit to form a valley, to "**collapse**" a valley to form a pit, to "**dimple**" a valley to form a saddle, to "**crumple**" a saddle to form a valley, to "**crease**" a saddle to form a ridge, to "**dent**" a ridge to form a saddle, to "**bulge**" a ridge to form a peak, and "**stretch**" a peak to form a ridge).
- Those shapes that do not make any transition - as they have no observable change in curvature (e.g. they are "**constant**").

This results in a total of **15 different deformation classes**, which we formalise as the *type of deformation:*

$$T \in [1, ..., 15] \tag{1}$$

that can occur over any given duration as defined by the relative change in the principal curvatures $\Delta\kappa_1$ and $\Delta\kappa_2$. These are shown formally in the extended Table 2 - indicating the transitions that can occur from each initial shape class (c.f. Table 1). As with the initial shape classes, in order to define the zero boundary region we employ a threshold term ($\Delta\kappa = 0 \iff -\theta_{change} < \Delta\kappa < \theta_{change}$).

We furthermore define the *extent of change* (E) to measure the degree to which this deformation occurs over the duration as based on the Koenderink "curvedness" value:

$$E = \sqrt{\frac{\Delta\kappa_1^2 + \Delta\kappa_2^2}{2}}. \tag{2}$$

3. Calculation

In the following sections, we illustrate how $\Delta\kappa_1$ and $\Delta\kappa_2$ can be calculated robustly. We assume here that incoming data forms a sequence of $2.5D$ depth-maps $Z(x, y, t)$ which can be re-projected into $\mathbb{R}^3$ - such that there is a neighbouring "patch" of associated data around any given point. Many other techniques rely on an underlying topology or triangulation to determine this neighbourhood - and can use the same underlying curvature estimation based on those representations (e.g. [7]). However, we instead directly exploit the array structure of the raw data depth-map.

3.1. Temporal Registration

We first compensate for global motion in intra-sequence frames. We assume here that any such motion is minimal, given that the data has been captured at a suitably fast frame rate, or has already been externally aligned. We wish to register on the basis of those rigid portions of the data, so as to highlight any deformation. In essence, we are seeking to address the issue of *local registration* in establishing the correspondence between points on the surface between frames (so that their curvature values may be compared).

We use the Iterative Closest Point algorithm [1] to minimise the Tukey M-estimator error distance between the two frames represented the data $Z(x, y, t)$ and $Z(x, y, t + 1)$ projected into $\mathbb{R}^3$. By using a Tukey operator we seek to only allow the best possible fitting for a subset to the data (we remove the 50% greatest distances outright) in order to reject outliers points caused by deformation. Having a fitting, we then project the final registered point back via the first frame camera for the resulting depth data alignment of frame $Z(x, y, t + 1)$.

3.2. Quadric Fitting

In order to calculate the curvature at every point, we desire an analytic description of the surface. For this purpose we fit a second order polynomial - a quadric - to the neighbourhood of each point. We simplify the process by seeking to align the quadric in the principal frame, at the origin with the normal aligned on the z-axis. This allows us to focus on fitting only an "extended" quadric (as opposed to the full 10 coefficient generalised form):

$$z = ax^2 + bxy + cy^2 + dx + ey. \tag{3}$$

We iteratively refine the fitting as the following [15]:

1. Take a patch of data around a point: $[X, Y, Z]$.
2. Translate the patch by its mean to the origin.
3. Fit a plane to the patch to derive initial normal.
4. Align the patch normal to the z-axis.
5. Perform a Least Squares fitting of the patch to the quadric: $\left[X^2, XY, Y^2, X, Y\right][a, b, c, d, e]^\top = z$.
6. Derive a new normal: $[-d, e, 1]^\top / (1 + d^2 + e^2)$.
7. Repeat from step 4 until normal aligned: $d = e = 0$.

A further advantage in using quadrics is that they can implicitly smooth through their minimisation to underlying data while fitting. Noise in the data is accommodated down to the level of scale at which true features can be fitted by the quadric. This can be further employed to effectively downsample particularly dense data, as dictated by the distance between sample points, and the size of the surrounding data patch.

3.3. Static Curvature Calculation

The calculation of curvature for the quadric at a point in frame $Z(x, y, t)$ can be derived from the first 3 coefficients of the fitted extended quadric (Eq. 3) at that point, such that:

$$\kappa_1 = (a + c) + ((a - c)^2 + b^2)^{1/2} \tag{4}$$

and

$$\kappa_2 = (a + c) - ((a - c)^2 + b^2)^{1/2}. \tag{5}$$

3.4. Temporal Curvature Change

Given a sequence of aligned frames at each time-step t for $\kappa_1(x, y, t)$ and $\kappa_2(x, y, t)$ then for a particular instance the temporal integration can be calculated for a window of duration d. Firstly by a convolution (denoted with $\star$) with a $1D$ Derivative of Gaussian (D_t) filter in the t dimension, then by spatial integration via convolution with Gaussian filters ($G_{x,y}$) in the x and y dimensions as follows:

$$\Delta\kappa_1(x, y, t) = G_x \star G_y \star D_t \star \kappa_1(x, y, t \pm d/2) \tag{6}$$

and similarly:

$$\Delta\kappa_2(x, y, t) = G_x \star G_y \star D_t \star \kappa_2(x, y, t \pm d/2). \tag{7}$$

For both convolutions we employ a Gaussian window with standard-deviation equal to 1. The window width d is the same both spatially and temporally over the period of integration. This is to accommodate larger motion which will potentially travel further and extend beyond the limit of the local registration. Ultimately, relying on convolution will lead to aliasing failure for motion that extends over longer intervals (and so requiring more advanced tracking).

4. Experimentation

Results of using the above calculations as the basis for determining the extent and type of deformation are presented. In all cases, the processing was performed "off-line" on an $3GHz$ Intel Pentium IV with $1GB$ of RAM, taking $5-60$ minutes to run in *Matlab* depending on down-sampling and quadric patch size. We set θ_{shape} and θ_{change} experimentally according to the scale of the data, and in order to best capture the range of deformation.

4.1. Synthetic Data

To numerically verify and illustrate our formulation, we first apply the calculation to synthetically generated data - as shown in Figure 3 below. Both sequences are generated over 5 frames at a resolution of 500×500, with no additional global motion. We down-sample and fit to this data at every 3^{rd} point with a 7×7 quadric (producing a final result of 166×166 - which is more efficient to process).

In sequence *(a)* the data is generated from a sinusoidal wave with increasing amplitude of 0.1 every frame. We set the thresholds $\theta_{shape} = 1.0$ and $\theta_{change} = 0.00001$ to effectively define all initial shapes as locally flat, and only verify the effects of altering the principal curvatures. The extent of change then shows the simultaneous increase in shape change toward the maximal bends in the surface. The types of deformation are also shown correctly to be primarily of bending and folding, with slight boundaries of buckling where the transition is made.

In sequence *(b)* we consider a $2D$ Gaussian with standard deviation $\sigma = 0.05$. This is scaled in z over the sequence by a factor of 0.1 between frames, using the same thresholds as above. The extent of change here nicely shows the differential increases expected, particularly large towards the apex of the data. For the types of deformation, the flat region round the peak is labelled as protruding, with folding then occurring as the relative angle between peak and base increases. The sides of the peak then suffer warping/bending as the peak protrudes further and increases the elongation of the surrounding curvatures.

4.2. Real Data

A dense stereo capture rig constructed from two cameras was calibrated and used in "burst" mode to capture a sequence from subjects making a number of basic expressions. These were sampled at 2.5 frames per second for a duration of 1.6 seconds - for a total of 4 images. Following this, dense stereo recovery was performed on each pair of simultaneous images to create a single high resolution 2048×3072 pixels frame of depth data. Additional masking of the data by removing any blue background, followed

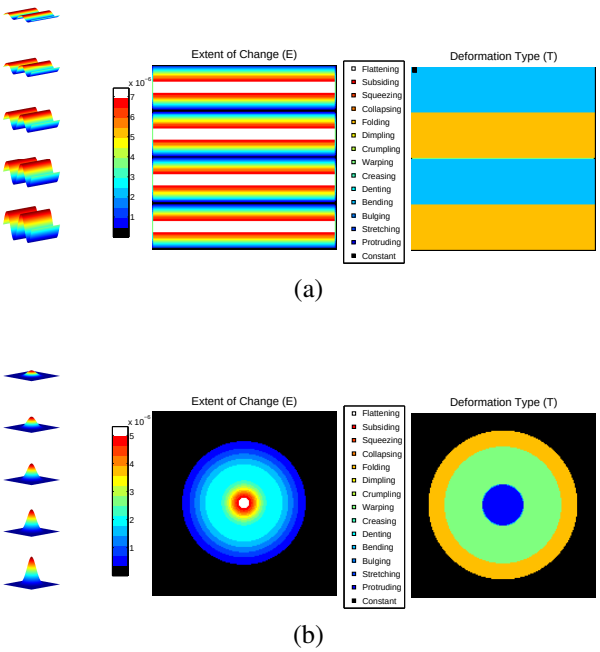


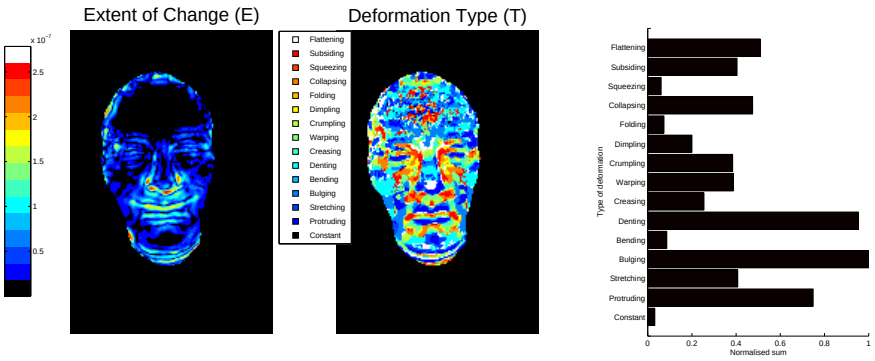
Figure 3. Synthetic data for expanding sinusoidal wave (a) and Gaussian peak (b).

by down-sampling to every 15^{th} point with a 121×121 neighbouring patch ($\approx 1cm^2$ in the original data) fitted to a quadric. We use thresholds $\theta_{shape} = 1.0 \times 10^{-9}$ and $\theta_{change} = 1.0 \times 10^{-5}$ to define the zero boundaries.

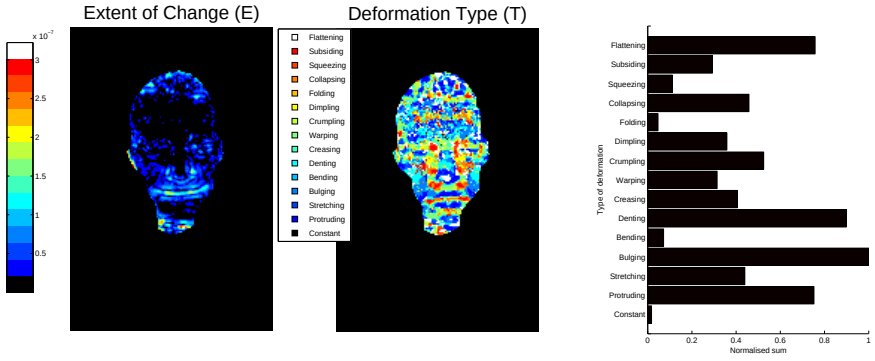
The results for extent and type as shown in Figure 4 reveal deformations that exemplify the areas of the face that alter for particular expressions. In the examples of *happiness* shown in (a) and (b), it is the regions round the mouth, nose and eyes that vary the most. This is most prominent in the relative extent of change, which clearly shows a smile forming. The distributions of types of deformation show a very similar pattern between subjects - as further revealed by comparing the histograms.

Similarly, in the examples of *surprise* as shown in (c) and (d), it is the brows which are seen to move in the characteristic upward direction, with ridges and valleys forming. Large portions of the face are also more constant, particularly around the cheeks. The extent of change again shows this most clearly, while the distribution of types indicates the occurrence of more bending and folding.

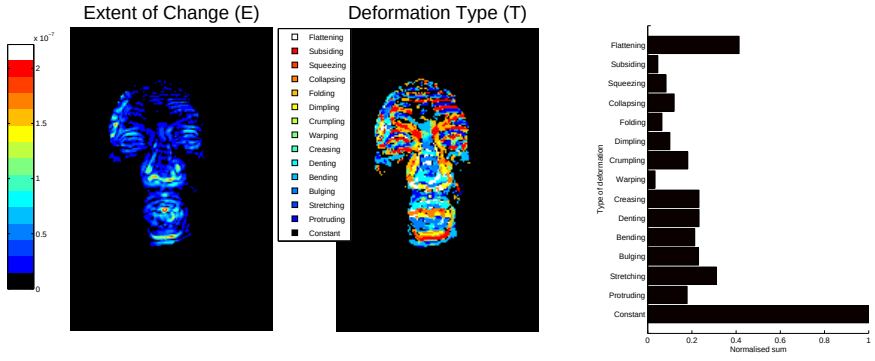
The Bhattacharyya distance [3] between the normalised histograms of (a) and (b) is 0.0875, and between (c) and (d) it is 0.2123. Conversely, between (a) to (c) it is 0.4230 and 0.4627 to (d). From (b) it is 0.4155 to (c) and 0.4852 to (d) respectively (0.0 = similar, 1.0 = dissimilar). This provides evidence that we are measuring consistent patterns of shape change for the two expressions.



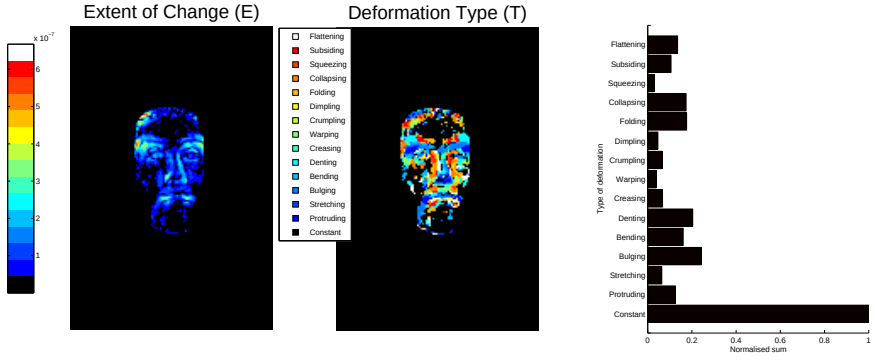
(a)



(b)



(c)



(d)

Figure 4. Real data extent and type of deformation for happiness (a)(b), and surprise (c)(d).

5. Conclusion

We have proposed a temporal extension of static curvature analysis, which uses initial and variations in the principle curvature values to characterise the type and extent of deformation ($T + E$). This serves to provide a simple and compact separation of 15 different classes. The control offered by the thresholding values allows us to specify the bearing that the initial shape and amount of change has on these categories.

However, while the results for real data are promising, they could be improved. Fundamentally, there is the crucial aspect of point-to-point surface correspondence between frames, especially over greater intervals of duration. Combining surface characteristics (flow, colour, texture, etc.) within a global constraining topology - could enable more robust tracking and local registration of points as they move and deform.

The inherent multi-resolution nature of complex surfaces could also be addressed by some form of hierarchical quadric fitting. Additionally, other types of rotational deformation (e.g. twisting and expanding) could be determined from the changes in the principal directions. This could be calculated within a more rigorous mathematical framework that takes advantage of the changes occurring in the first and second fundamental forms.

In general, this promises to be a useful technique that can be employed when analysing data captured from novel video-rate $4D$ systems at higher-frame rates (with a suitable lag for integration). This would still be performed "off-line" (as per the example results above), but could be substantially faster using optimised, compiled versions of the algorithm. We hope to improve and build on this work to further investigate the automatic classification of faces along the variational axes of both form and intensity, with the possible identification of particular expressions using a statistical (eigenfaces) approach.

6. Acknowledgements

This work is supported by a UK EPSRC Imaging Faraday CASE award in conjunction with Dimensional Imaging Ltd. (`www.di3d.com`).

References

[1] P. Besl and N. McKay. A method for registration of 3D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[2] P. J. Besl and R. C. Jain. Invariant surface characteristics for 3D object recognition in range images. *Computer Vision, Graphics, and Image Processing*, 33(1):33–80, 1986.

[3] A. Bhattacharya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of Calcutta Maths Society*, 35:99–110, 1943.

[4] F. L. Bookstein. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, 1991.

[5] D. M. Cash, T. K. Sinha, C. Chen, B. M. Dawant, W. C. Chapman, M. I. Miga, and R. L. Galloway. Identification of deformation using invariant surface information. In *Proc. SPIE*, volume 5367, pages 140–150, 2004.

[6] M. K. Chung, K. J. Worsley, T. Paus, C. Cherif, D. L. Collins, J. N. Giedd, J. L. Rapoport, and A. C. Evans. A unified statistical approach to deformation-based morphometry. *NeuroImage*, 14(3):595–606, 2001.

[7] R. V. Garimella and B. K. Swartz. Curvature estimation for unstructured triangulations of surfaces. Technical Report LA-UR-03-8240, Los Alamos National Laboratory, Nov 2003.

[8] J. J. Koenderink. *Solid Shape*. MIT Press, Cambridge, 1990.

[9] J. J. Koenderink and A. J. van Doorn. Optic flow. *Vision Research*, 26(1):161–180, 1986.

[10] J. J. Koenderink and A. J. van Doorn. Surface shape and curvature scales. *Image and Vision Computing*, 10(8):557–565, 1992.

[11] X. Lu, D. Colbry, and A. K. Jain. Matching 2.5D scans for face recognition. In *Proc. of the International Conference on Biometric Authentication*, pages 30–36, 2004.

[12] X. Lu and A. K. Jain. Deformation analysis for 3D face matching. In *Proc. 7th IEEE Workshop on Applications of Computer Vision*, pages 99–104, 2005.

[13] T. C. Lukins and R. B. Fisher. Colour constrained 4D flow. In *Proceedings of the British Machine Vision Conference*, pages 340–348, 2005.

[14] T. McInerney and D. Terzopoulos. A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4D image analysis. *Computerized Medical Imaging and Graphics*, 19(1):69–83, 1995.

[15] A. M. McIvor and R. J. Valkenburg. Principal frame and principal quadric estimation. *Image and Vision Computing New Zealand*, pages 55–60, 1996.

[16] H. Spies, B. Jähne, and J. L. Barron. Range flow estimation. *Computer Vision Image Understanding*, 85(3):209–231, 2002.

[17] Y. Wang, M. Gupta, S. Zhang, S. Wang, X. Gu, D. Samaras, and P. Huang. High resolution tracking of non-rigid 3D motion of densely sampled data using harmonic maps. In *Proc. International Conference on Computer Vision*, volume 1, pages 388–395, 2005.

[18] I. A. Ypsilos, A. Hilton, and S. Rowe. Video-Rate capture of dynamic face shape and appearance. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 117–122, 2004.

[19] L. Zhang, N. Snavely, B. Curless, and S. M. Seitz. Spacetime faces: High-resolution capture for modeling and animation. In *ACM Annual Conference on Computer Graphics*, pages 548–558, August 2004.

[20] M. Zhili, J. Siebert, W. Cockshott, and A. Ayoub. Constructing dense correspondences to analyze 3D facial change. In *Proc. 17th International Conference on Pattern Recognition*, volume 3, pages 144–148, 2004.