

Enhanced Real-Time Face Models from Stereo Imaging for Gaming Applications

Mircea C. Ionita, Istvan Andorko and Peter Corcoran, *Senior Member IEEE*
College of Engineering & Informatics, National University of Ireland Galway, Galway, Ireland

Abstract—Techniques for improved 2D active appearance face models are presented. When these are applied to stereoscopic image pairs we show that sufficient information on image depth is obtained to generate an approximate 3D face model. Two techniques are investigated, the first based on 2D+3D AAMs and the second using methods based on thin plate splines. The resulting 3D models can offer a practical real-time face model which is suitable for a range of applications in computer gaming. Due to the compact nature of AAMs these are also very suitable for use in embedded devices such as gaming peripherals.

Keywords – AAM; face models; stereo imaging; game applications; user interfaces.

I. INTRODUCTION

Face detection and tracking technology has become commonplace in digital cameras in the last year or so. All of the practical embodiments of this technology are based on Haar classifiers and follow some variant of the classifier cascade originally proposed by Viola and Jones [8]. These Haar classifiers are rectangular and by computing a grayscale integral image mapping of the original image it is possible to implement a highly efficient multi-classifier cascade. These techniques are also well suited for hardware implementations [23].

Now, despite the rapid adoption of such in-camera face tracking, the tangible benefits are primarily in improved enhancement of the global image. An analysis of the face regions in an image enables improved exposure and focal settings to be achieved. However current techniques can only determine the approximate face region and do not permit any detailed matching to facial orientation or pose. Neither do they permit matching to local features within the face region. Matching to such detailed characteristics of a face region would enable more sophisticated use of face data and the creation of real-time facial animations for use in, for example, gaming avatars. Another field of application for next-generation gaming technology would be the use of real-time face models for novel user interfaces employing face data to initiate game events, or to modify difficult levels based on the facial expression of a gamer.

In this paper we investigate a particular class of 2D affine models, known as active appearance models (AAM), which are relatively fast and are sufficiently optimal to be suitable for in-camera implementations. To improve the speed and robustness of these models we have investigated several enhancements, in particular we describe improvements to (i) deal with directional lighting effects and (ii) make use of the full color range to

improve accuracy and convergence of model to a detected face region.

Additionally we present preliminary investigations into the use of stereo imaging to improve model registration by using two real-time video images with slight variations in spatial perspective. As AAM models are essentially 2D affine models the use of a real-time stereo video stream opens interesting possibilities to create a full 3D face model from the 2D real-time models we currently employ.

In section II we do an overview of these models and describe the principle steps in constructing simple AAM models; in section III we focus on enhancement to handle directional lighting and present some results and examples; we also investigate the use of the full color range in such models and demonstrate that color information can be used advantageously to improve both the accuracy and speed of convergence of the model; we discuss here as well results for one example usage of such models, a method for performing face recognition; comparative analysis shows that results from our improved model are significantly better than those obtained from a conventional AAM or from the well-known *eigenfaces* method for performing face recognition; in section IV we present a differential stereo model which can be used to further enhance model registration and which offers the means to extend our 2D real-time models to a pseudo 3D model; in section V we present an approach for generating realistic 3D avatars based on a computationally reduced thin plate spline warping technique; the method incorporates modeling enhancements described in sections III and IV; finally, in section VI we summarize our conclusions and provide a further discussion of the potential for use of AAM models across a range of gaming applications.

II. AAM OVERVIEW

This section explains the fundamentals of creating a statistical model of appearance and of fitting said model to image regions.

A. Statistical Models of Appearance

AAM was proposed by Cootes et al. [1] in 1998 as a deformable model, capable of interpreting and synthesizing new images of the object of interest. Statistical Models of Appearance represent both the shape and texture variations and the correlations between them for a particular class of objects. Example members of the class of objects to be modeled are annotated by a number of landmark points. The shape is defined by the number of landmarks chosen to best depict the contour of the object of interest, in our case a person's face.

B. Model Fitting to an Image Region

After a statistical model of appearance is created, an AAM algorithm can be employed to fit the model to a new, unseen, image region. The statistical model is linear in both shape and texture. However, fitting the model to a new image region is a non-linear optimization process. The fitting algorithm works by minimizing the error between a query image and the equivalent model-synthesized image.

In this paper we use an optimization scheme which is robust to directional variations in illumination. This relies on the fact that lighting information is decoupled from facial identity information. This can be seen as an adaptation of the method proposed by Batur *et al* [2]. These authors use an adaptive gradient where the gradient matrix is linearly adapted according to the texture composition of the target image, generating an improved estimate of the actual gradient. In our model the separation of texture into lighting dependent and lighting independent subspaces enables a faster adaptation of the gradient.

C. Initialization of the Model within an Image

Prior to implementing that AAM fitting procedure it is necessary to initialize the model within an image. To detect faces we employ a modified Viola-Jones face detector [8] which can accurately estimate the position of the eye regions within a face region. Using the separation of the eye regions also provides an initial size estimate for the model fitting. The speed and accuracy of this detector enables us to apply the AAM model to large unconstrained image sets without a need to pre-filter or crop face regions from the input image set.

III. MODEL ENHANCEMENTS – ILLUMINATION AND MULTI-CHANNEL COLOUR REGISTRATION

A. Building an Initial Identity Model

The reference shape used to generate the texture vectors should be the same one for all models, i.e. either identity or directional lighting models. Our goal is to determine specialized subspaces, such as the identity subspace or the directional lighting subspace.

We first need to model only the identity variation between individuals. For training this identity-specific model we only use images without directional lighting variation. Ideally these face images should be obtained in diffuse lighting conditions. Textures are extracted by projecting the pixel intensities across the facial region, as defined by manual annotation (see Fig. 1) into the reference shape – chosen as the mean shape of the training data.

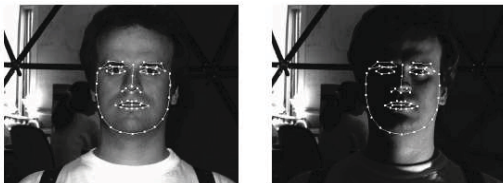


Fig. 1: Example of annotations used for Yale B Database

The number of landmark points used should be kept fixed over the training data set. In addition to this, each landmark point must have the same face geometry correspondence for

all images. The landmarks should predominantly target facial fiducial points, which permit a good description of facial geometry, allowing as well the extraction of geometrical differences between different individuals. The facial textures corresponding to images of individuals in the Yale database with frontal illumination are represented in Fig. 2(a).



(a)



(b)



(c)

Fig. 2: (a) Variation between individuals; (b) estimated albedo of the individuals; (c) albedo eigen-textures with 95% energy preservation

The identity model can now be generated from the *albedo* images based on the standard PCA technique.

B. Building a Model for Directional Lighting Variations

Consider now all facial texture which exhibit directional lighting variations from all 4 subsets. These texture are firstly projected onto the previously built subspace of individual variation, ULS. We note these texture vectors, containing some directional lighting information, with g .

Note that g contains both identity and directional lighting information. The same reference shape is used to obtain the new texture vectors g , which ensures that the previous and new texture vectors have all equal lengths. In Fig. 3(a) a random selection of faces is shown. The projection of the texture vectors g onto ULS gives the sets of optimal texture parameter vectors as in:

$$b_{ident}^{(opt)} = \Phi_{ident}^T (g - \bar{t}) \quad (1)$$

The back-projection stage returns the texture vector, optimally synthesized by the identity model. The projection/back-projection process filters out all the variations which could not be explained by the identity model. Thus, for this case, all directional lighting variations are filtered out by this process,

$$g_{filt} = \bar{t} + \Phi_{ident} b_{ident}^{(opt)} \quad (2)$$

Continuing with the procedure for the examples in Fig. 3(a), their filtered versions are shown in Fig. 3(b).



Fig. 3(a): A reference sample subset of images with various directional lighting effects

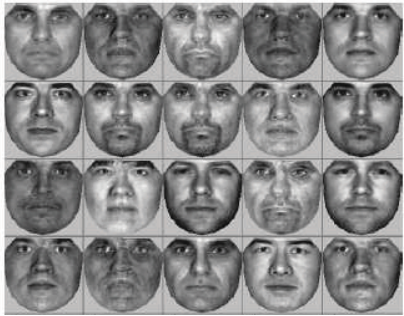


Fig. 3(b): Face samples from 3(a) with the contribution of directional lighting removed by filtering (eq. 5)

The residual texture is further obtained as the difference between the original texture and the synthesized texture which retained only the identity information. This residual texture normally retains the information other than identity.

$$t_{res} = g - g_{filt} = g - \bar{t} - \Phi_{ident} b_{ident}^{(opt)} \quad (3)$$

The residual images give the directional lighting information - Fig. 4. These residuals are then modeled using PCA in order to generate the directional lighting subspace.

C. Creating a Merged Face Model

As described above, three separate components of the face model have been generated. These are: (i) the shape model of the face, (ii) texture model encoding identity information, and (iii) the texture model for directional lighting. The resulting texture subspaces are also orthogonal due to the approach described above. The fusion between the two texture models can be realized by a weighted concatenation of parameters:

$$c = \begin{bmatrix} W_s b_s \\ b_{ident} \\ W_{light} b_{light} \end{bmatrix} \quad (4)$$

where W_{light} and W_s are two vectors of weights used to compensate for the differences in units between the two sets of

texture parameters, and for the differences in units between shape and texture parameters, respectively.



Fig. 4: Images of Fig 3(b) subtracted from images of Fig 3(a) to yield a set of difference (residual) images.

D. Fitting the Lighting Enhanced Model

The conventional AAM algorithm uses a gradient estimate built from training images and thus cannot be successfully applied to images where there are significant variations in illumination conditions. The solution proposed by Batur et al. is based on using an adaptive gradient AAM [6]. The gradient matrix is linearly adapted according to texture composition of the target image. We further modify the approach of Batur [2] to handle our combined ULS and DLS texture subspace. The derivation and justification of this approach is quite complex and so the interested reader is referred to [3], [4] for full details.

E. Colour Space Enhancements

When a typical multi-channel image is represented in a conventional color space such as RGB, there are correlations between its channels. For natural images, the cross-correlation coefficient between B and R channels is ~ 0.78 , between R and G channels is ~ 0.98 , and for G and B channels is ~ 0.94 [24]. This inter-channel correlation explains why previous authors [25] obtained poor results using RGB AAM models.

Ohta's space [10] realizes a statistically optimal minimization of the inter-channel correlations, i.e. decorrelation of the color components, for natural images. The conversion from RGB to I1I2I3 is given by the simple linear transformations in (5a-c).

$$I_1 = \frac{R+G+B}{3} \quad (5a)$$

$$I_2 = \frac{R-B}{2} \quad (5b)$$

$$I_3 = \frac{2G-R-B}{4} \quad (5c)$$

I1 represents the achromatic (intensity) component, while I2 and I3 are the chromatic components. By using Ohta's space the AAM search algorithm becomes more robust to variations in lighting levels and color distributions. We present a summary of comparative results across different color spaces in Tables I and II. The interested reader is referred to [38] and [3] for a more detailed analysis and description of this work.

TABLE I
TEXTURE NORMALISATION RESULTS ON (PIE) SUBSET 2 (*Unseen*)

Model	Success [%]	Pt-Crv (Mean/Std)	Pt-Pt (Mean/Std)	PCTE (Mean/Std)
Greyscale	88.46	3.93/2.00	6.91/5.45	-
RGB GN	80.77	3.75/1.77	7.09/4.99	7.20/2.25
CIELAB GN	100	2.70/0.93	4.36/1.63	5.91/1.19
I1I2I3 SChN	100	2.60/0.93	4.20/1.45	5.87/1.20
RGB SChN	73.08	4.50/2.77	8.73/7.20	7.25/2.67
CIELAB SChN	88.46	3.51/2.91	6.70/8.29	6.28/2.09
I1I2I3 GN	92.31	3.23/1.21	5.55/2.72	6.58/1.62

TABLE II
CONVERGENCE RESULTS ON *Unseen* DATABASES

Model	Success Rate [%]	Pt-Crv (Mean/Std/Median)	PTE (Mean/Std/Median)
db1-Grayscale*	92.17	5.10 1.66 4.90	4.28 1.03 4.21
db1-RGB-none	99.13	4.94 1.37 4.82	10.09 1.58 9.93
db1-RGB-G	98.26	4.98 1.44 4.65	7.49 1.98 7.02
db1-RGB-Ch	87.83	5.32 1.65 5.08	6.33 1.40 5.95
db1-I1I2I3-Ch	99.13	3.60 1.32 3.32	5.10 1.01 4.85
db1-I1I2-Ch	99.13	4.25 1.65 3.79	8.26 4.11 6.10
db2-Grayscale*	75.73	4.17 1.44 3.67	5.12 4.24 4.03
db2-RGB-none	84.47	4.02 1.40 3.69	12.43 3.43 12.41
db2-RGB-G	94.17	3.74 1.45 3.23	9.04 1.83 8.97
db2-RGB-Ch	62.14	4.01 1.60 3.46	7.70 4.26 6.06
db2-I1I2I3-Ch	88.35	3.31 1.26 2.98	6.16 2.28 5.73
db2-I1I2-Ch	87.38	3.60 1.55 3.04	10.00 3.41 8.94
db3-Grayscale*	63.89	4.85 2.12 4.26	4.90 3.44 3.98
db3-RGB-none	72.22	4.44 1.79 3.99	14.23 4.79 13.34
db3-RGB-G	65.28	4.55 2.03 4.01	9.68 2.81 9.27
db3-RGB-Ch	59.72	5.02 2.04 4.26	7.16 4.91 5.74
db3-I1I2I3-Ch	86.81	3.53 1.49 3.15	6.04 2.56 5.20
db3-I1I2-Ch	86.81	3.90 1.66 3.41	6.60 1.94 6.30

F. Uses of Statistical Models and AAM

In section V we will describe the use of an improved AAM model for use in face recognition. However there are a multitude of alternative applications for such models. These models have been widely used for face tracking [7], and measuring facial pose and orientation.

In other research we have demonstrated the use of AAM models for detecting phenomena such as eye-blink [16], analysis and characterization of mouth regions [17] and facial expressions [18]. In such context these models are more sophisticated than other pattern recognition methods which can only determine if, for example, an eye is in an open or closed state. Our models can determine other metrics such as the degree to which an eye region is open or closed or the gaze direction of the eye [16]. This opens the potential for sophisticated game avatars or novel gaming UI methods

G. Building a Combined Model

A notable applicability of the directional lighting sub-model, generated from a grayscale training database, is that it can be efficiently incorporated into a color face model. This process is illustrated in Fig. 5. The left-hand process diagram illustrates the partitioning of the model texture space into orthogonal ULS and DLS subspaces. The right-hand side process diagram shows how the DLS subspace can be used to train a color ULS, implemented in the Otha color space. This yields a full color ULS which retains the orthogonality with the DLS and when combined with it yields an enhanced AAM model incorporating shape + DLS + color ULS subspaces. The color ULS has the same improved fitting characteristics as the

color model [38] and so this combined model exhibits both improved registration and robustness to directional lighting.

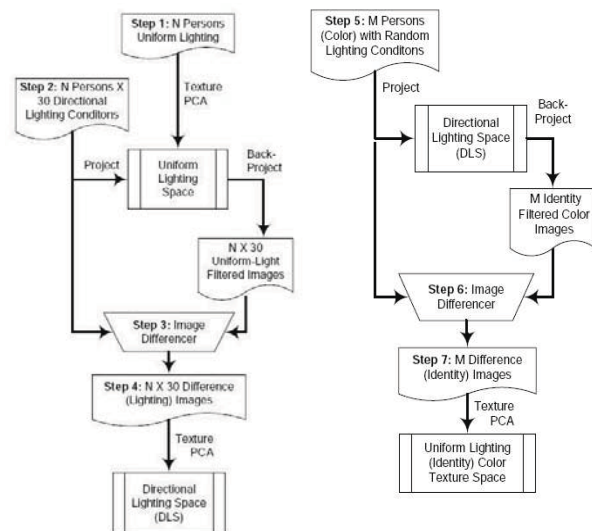


Fig 5: Process steps to build a color extension of the combined DLS + ULS model for face recognition.

H. Model Application to Face Recognition

1) Benchmarking for AAM-based Face Recognition.

The recognition tests which follow have been performed by considering the large gallery test performance, as proposed in [19]. As a benchmark with other methods we decided to compare relative performance with respect to the well-known eigenfaces method [20]. Detailed results of these tests are reported in [3]. There is a reported modest improvement of 5%-8% to be achieved in using a color AAM method (RGB) over a grayscale AAM. The performance of the color AAM is approximately equal to that of both grayscale and color eigenfaces methods.

2) Tests on the Improved AAM Model.

Again a more detailed report of these tests is given in [3]. We note that none of the color AAM techniques based on RGB color space can compete with the conventional eigenface method of face recognition. Conversely, all of the I1I2I3 based models perform at least as well as the eigenface method, even when the model has been trained on a different database. When trained on the same database we conclude that the I1I2I3 SChN model outperforms the eigenface method by at least 10% when the first 50 components are used. If we restrict our model to the first 5 or 10 components then the differential is about 20% in favor of the improved AAM model.

IV. MODEL ENHANCEMENTS – DIFFERENTIAL AAM FROM REAL-TIME STEREO CHANNELS

A. Hardware Architecture of Stereo Imaging System

The general architecture of the system is shown in Fig 6 below. The two CMOS sensors are connected to an FPGA which incorporates a PowerPC core and associated SDRAM. Additional system components have been added to implement a dual stereo image processing pipeline [37].

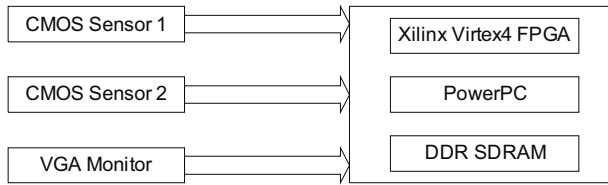


Fig. 6: General architecture

The development board is a Xilinx ML405 development board, with a Virtex 4 FPGA, a 64 MB DDR SDRAM memory, and a PowerPC RISC processor. The clock frequency of the system is 100 MHz. The internal architecture of the system is shown in Fig. 7. The sensor used is a 1/3 inch SXGA CMOS sensor made by Micron. It has an active zone of 1280x1024 pixels. It is programmable through the I2C interface. It works at 13.9 fps and the clock frequency is 25 MHz. This sensor was selected because of its small size, low cost and the specifications of these sensors are satisfactory for this project. The interested reader is referred to [37] for additional details and a more detailed review of stereo imaging.

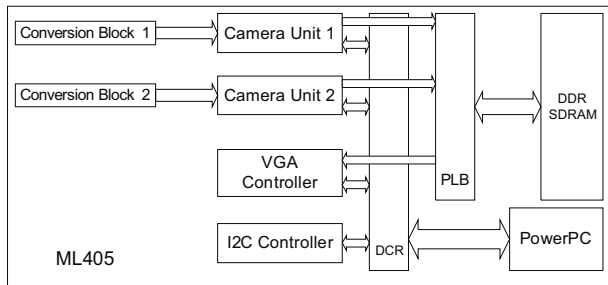


Fig. 7: Internal architecture

This system enables real-time stereo video capture with a fixed distance between the two imaging sensors.



Fig. 8: Stereo face image pair example

B. Determination of a Depth Map

When using two sensors for stereo imaging, the problem of parallax effect appears. Parallax is an apparent displacement or difference of orientation of an object viewed along two different lines of sight, and is measured by the angle or semi-angle of inclination between those two lines.

The advantage of the parallax effect is that with the help of this, depth maps can be computed. The algorithm requires pairs of rectified images, [12] which means that corresponding epipolar lines are horizontal and on the same height. The search of corresponding pictures takes place in horizontal direction only. For every pixel in the left image the goal is to find the corresponding pixel in the right image.

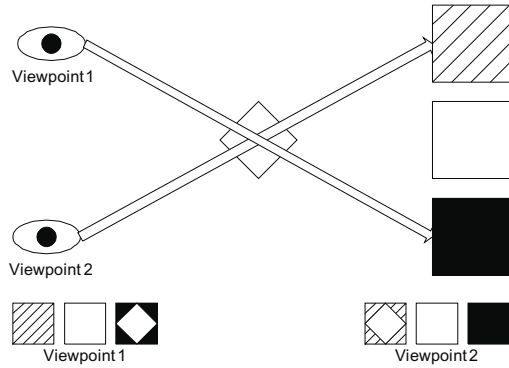


Fig 9: The Parallax Effect

It is practically impossible to find corresponding single pixels, thus we have used windows of different sizes (3x3; 5x5; 7x7). The size of window is computed based on the value of the local variation of each pixel. [13] The formula for the computation of the local variation is that one proposed in [13] and is shown below:

$$LV(p) = \sum_{i=1}^N \sum_{j=1}^N |I(i,j) - \mu| \quad (6)$$

where μ is the average grayscale value of image window, and N is the selected square window size.

The first local variation calculation is made over a 3x3 window. After this, the points with a value under a certain threshold are marked for further processing. The same operation is done for 5x5 and 7x7 windows as well. The sizes of the windows is stored for use in the depth map computation. The operation to compute the depth map is the Sum of Absolute Differences for RGB images (SAD). The value of SAD is computed for up to a maximum value of d on the x line. After all the SAD values have been computed, the minimum value of $SAD(x,y,d)$ is chosen, and the value of d from this minimum will be the value of the pixel in the depth map. [12] At searching the minimum, there are some problems that we should be aware of. If the minimum is not unique, or its position is d_{\min} or d_{\max} , the value is discarded. Instead of just seeking the minimum, it is helpful to track the three smallest SAD values as well. The minimum defines a threshold above which the third smallest value must lie. Otherwise, the value is discarded.

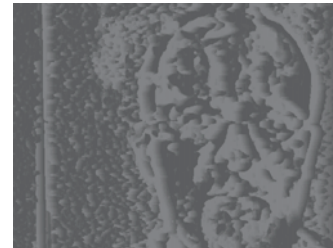


Fig. 10: Depth map result for the stereo image pair in Fig. 8

One of conditions for our depth map computation algorithm to work properly is that the stereo image pairs should contain strong contrast between the colors within the image and there should not be large areas of nearly uniform color. Other

researchers who attempted the implementation of this algorithm used computer generated stereo image pairs, which contained multiple colors [13], [26]. In our case, the results after applying the algorithm for faces were sub-optimal because the color of facial skin is uniform across most of the face region and the algorithm was not able to find exactly similar pixels in the stereo image pair.

C. AAM Enhanced Shape Model

In the earlier discussion of section III we presented a face model with two, orthogonal texture spaces. In this section we will present the development of a dual orthogonal shape subspace which is derived from the difference and averaged values of the landmark points derived from the right-hand and left hand stereo face images. This separation provides us with an improved 2D registration estimate from the averaged landmark point locations and an orthogonal subspace derived from the different values.

This second subspace enables an improved determination of the SAD values and the estimation of an enhanced 3D surface view over the face region.



Fig. 11: Fitted AAM face model on the stereo pair in Fig. 9

An example of fitting the model on a stereo image pair is given in Fig. 11, showing the identified positions of the considered landmarks. The corresponding triangulated shapes are then shown in Fig. 12. The landmarks are used as control points for generating the 3D shape, based of their relative 2D displacement in the two images. The result is shown in Fig. 13.

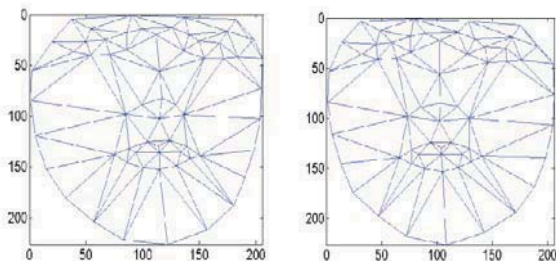


Fig. 12: Corresponding triangulated meshes for fitted model in Fig. 12

The 3D shape model allows for 3D constraints to be imposed, making the face model more robust to pose variations; it also reduces the possibility of generating unnatural shape instances during the fitting process, subsequently reducing the risk of an erroneous convergence. Efficient fitting algorithms for the new, so called 2D+3D, model have already been developed [27], [28], [29].

Full 3D face models, called 3D morphable models (3DMM), have also been proposed [31]. Yet, these models

have a high complexity and significant computational requirements thus we prefer the approaches based on the simpler AAM techniques. By incorporating the enhancements outlined in section III we feel they offer a more useful approach for implementation in embedded systems.

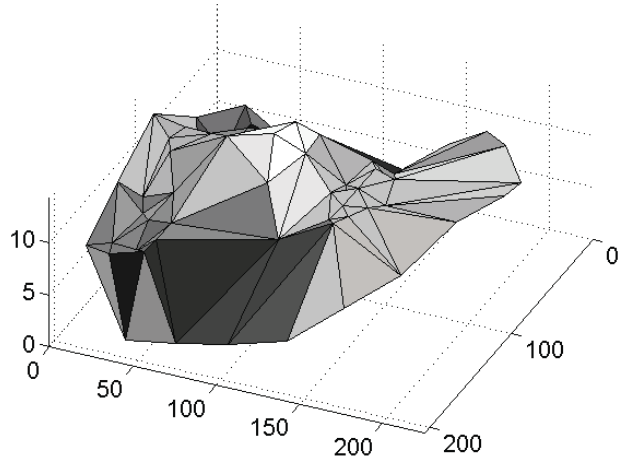


Fig. 13: 3D shape from 2D stereo data with triangulation-based warping

V. MODEL APPLICATION – 3D GAMING AVATARS

In previous sections we employed a triangulation-based, piecewise affine method for generating and fitting our statistical face models; it is commonly preferred due to its reduced computational costs. Above we used the Delaunay triangulation for partitioning the convex hull of our control points. The points inside triangles are mapped via an affine transformation which uniquely assigns the corners of a triangle to their new positions.

A different warping method, that yields a denser 3D representation, is based on thin plate splines (TPS), which have been originally introduced in [33]. For a more detailed discussion on the use of TPS for improving the convergence accuracy of color AAMs refer to [32]. The possibility of using TPS-based warping for estimating 3D face profiles has been shown in [3], [32].

In the context of generating realistic 3D avatars, the choice of TPS-based warping technique offers a more suitable solution. This technique is more complex than the piecewise linear warping employed above; yet, note that important progress has been made to simplify and reduce its initial computational complexity. A more detailed discussion on this aspect can be found in [3]. We summarize next the main steps for generating the TPS-based warping. TPS-based warping represents a nonrigid registration method, built upon an analogy with a theory in mechanics. Namely, the analogy is made with minimizing the bending energy of a thin metal plate on which pressure is exerted using some point constraints. The bending energy is then given by a quadratic form; the spline is represented as a linear combination (superposition) of eigenvectors of the bending energy matrix:

$$f(x, y) = a_1 + a_x x + a_y y + \sum_{i=1}^p w_i U(\|(x_i, y_i) - (x, y)\|). \quad (7)$$

where $U(r) = r^2 \log(r)$; (x_i, y_i) are the initial control points. $\mathbf{a} = (a_1 \ a_x \ a_y)$ defines the affine part, while w defines the nonlinear part of the deformation. The total bending energy is expressed as

$$I_f = \iint_{\mathbb{R}^2} \left(\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right) dx dy, \quad (8)$$

The surface is deformed such that to have minimum bending energy. The conditions that need to be met so that (7) is valid, i.e., so that $f(x, y)$ has second-order derivatives, are given by

$$\sum_{i=1}^p w_i = 0 \quad (9)$$

and

$$\sum_{i=1}^p w_i x_i = 0; \quad \sum_{i=1}^p w_i y_i = 0. \quad (10)$$

Adding to this the interpolation conditions $f(x_i, y_i) = v_i$, (7) can now be written as the linear system in (10):

$$\begin{bmatrix} K & P \\ P^T & O \end{bmatrix} \begin{bmatrix} \mathbf{w} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{v} \\ \mathbf{o} \end{bmatrix}, \quad (11)$$

where $K_{ij} = U(\|(x_i, y_i) - (x_j, y_j)\|)$, O is a 3×3 matrix of zeros, \mathbf{o} is a 3×1 vector of zeros, $P_{ij} = (1, x_i, y_i)$; \mathbf{w} and \mathbf{v} are the column vectors formed by w_i and v_i , respectively, while $\mathbf{a} = [a_1 \ a_x \ a_y]^T$.

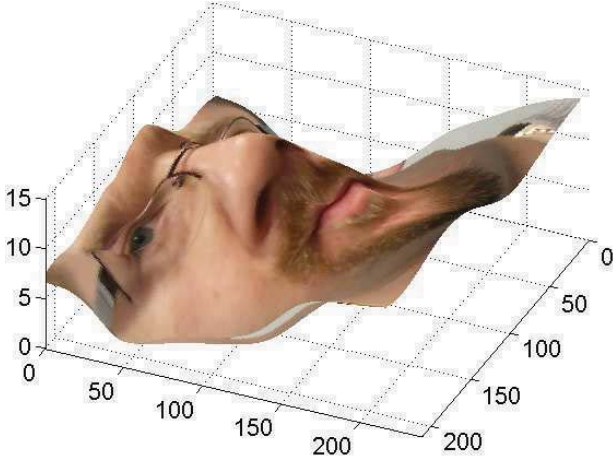


Fig. 14: Estimated 3D profile from 2D stereo data using Thin Plate Spline-based warping

The main drawback of using the thin plate splines was their high computational load. The solution requires the inversion of a $p \times p$ matrix (the bending energy matrix) which has a computational complexity of $O(N^3)$, where p is the number of points in the dataset (i.e., the number of pixels in the image); furthermore, the evaluation process is $O(N^2)$. Since their introduction though, important progress has been made in order to speed this process up. An approximation approach was proved in [34] to be very efficient in dealing with the first problem, reducing greatly the computational burden. As far as the evaluation process is concerned, the multilevel fast

multipole method (MLFMM) framework was described in [35] for the evaluation of two-dimensional polyharmonic splines, while in [36] this work was extended for the specific case of TPS, showing that a reduction of the computational complexity from $O(N^2)$ to $O(N \log N)$ is indeed possible. Thus the computational difficulties involving the use of TPS have been greatly reduced.

Based on this warping technique we are able to generate 3D facial profiles as shown in Fig. 14.

VI. CONCLUSIONS

We have presented methods to build improved AAM facial models which condense significant information about facial regions within a relatively small data model. Methods have been described which allow models to be constructed with orthogonal texture and shape subspaces. These allow compensation for directional lighting effects and improved model registration using color information.

These improved models may then be applied to stereo image pairs to deduce 3D facial depth data. This enables the extension of the AAM to provide a 3D face model. Here we presented preliminary results of two approaches – one based on 2D + 3D AAM and a second approach based on *thin plate spline* warpings. Our results are encouraging, particularly those based on *thin plate splines* which are shown to produce an acceptable 3D rendering of the face data.

Our results, although still preliminary, suggest that these extended AAM based techniques when combined with stereoscopic image data offer potential for improved user interface methods and the generation of dynamic real-time avatars for computer gaming applications.

ACKNOWLEDGMENT

This project has been part-funded by the Irish Research Council for Science, Engineering and Technology (IRCSET) and Tessera (Ireland) Ltd. .

REFERENCES

- [1] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models", *Lecture Notes in Computer Science*, vol. 1407, pp. 484–, 1998.
- [2] A. U. Batur and M. H. Hayes, "Adaptive active appearance models," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1707–1721, 2005.
- [3] M. Ionita, "Advances in the design of statistical face modeling techniques for face recognition", *PhD Thesis*, NUI Galway, 2009.
- [4] M. Ionita and P. Corcoran, "A Lighting Enhanced Facial Model: Training and Fast Optimization Scheme", submitted to *Pattern Recognition*, May 2009.
- [5] M. Ionita, I. Bacivarov, and P. Corcoran, "Separating directional lighting variability in statistical face modelling based on texture space decomposition," in *Proceedings of the 15th International Conference on Digital Signal Processing (DSP'07)*, Cardiff, Wales, UK, July 2007, pp. 252–255.
- [6] F. Kahraman, M. Gokmen, S. Darkner, and R. Larsen, "An active illumination and appearance (AIA) model for face alignment," *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pp. 1–7, June 2007.
- [7] P. Corcoran, M.C. Ionita, I. Barcivarov, "Next generation face tracking technology using AAM techniques," *Signals, Circuits and Systems, 2007. ISSCS 2007. International Symposium on*, Volume 1, p 1-4, 13-14 July 2007.

- [8] P. A. Viola, M. J. Jones, "Robust real-time face detection", *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [9] J. J. Gerbrands, "On the relationships between SVD, KLT and PCA." *Pattern Recognition*, vol. 14, no. 1-6, pp. 375–381, 1981.
- [10] Y. Ohta, T. Kanade, and T. Sakai, "Color Information for Region Segmentation", *Ó Comput. Graphics Image Process.*, vol. 13, pp. 222–240, 1980.
- [11] M. B. Stegmann and R. Larsen, "Multi-band modelling of appearance," *Image and Vision Computing*, vol. 21, no. 1, pp. 61–67, jan 2003. [Online]. Available: <http://www2.imm.dtu.dk/pubdb/p.php?1421>
- [12] K. Muhlmann, D. Maier, J. Hesser, R. Manner, "Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation", *International Journal of Computer Vision*, vol. 47, numbers 1-3, pp. 79 – 88, April 2002.
- [13] C. Georgoulas, L. Kotoulas, G. Ch. Sirakoulis, I. Andreadis, A. Gasteratos, "Real-Time Disparity Map Computation Module", *Microprocessors and Microsystems* 32, pp. 159 – 170, 2008.
- [14] Y. Li, W. Ito, *US Patent Application* 20070071347.
- [15] H. Kameyama, W. Ito, *US Patent Application* 20060257047.
- [16] I. Bacivarov, M. Ionita, P. Corcoran, "Statistical Models of Appearance for Eye Tracking and Eye-Blink Detection and Measurement". *IEEE Transactions on Consumer Electronics*, August 2008.
- [17] I. Bacivarov, M.C. Ionita, and P. Corcoran, A Combined Approach to Feature Extraction for Mouth Characterization and Tracking, in *Signals and Systems Conference, 2008. (ISSC 2008). IET Irish*, Volume 1, p 156–161, Galway, Ireland 18-19 June 2008.
- [18] J. Shi, A. Samal, and D. Marx, "How effective are landmarks and their geometry for face recognition?" *Comput. Vis. Image Underst.*, vol. 102, no. 2, pp. 117–133, 2006.
- [19] P. J. Phillips, P. Rauss, and S. Der, "FERET recognition algorithm development and test report," U.S. Army Research Laboratory, *Tech. Rep.*, 1996.
- [20] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR '91)*, 586–591, 1991.
- [21] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [22] D. Gonzalez-Jimenez and J. Alba-Castro, "Toward pose-invariant 2-d face recognition through point distribution models and facial symmetry," *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 3, pp. 413–429, Sept. 2007.
- [23] A. Bigdeli, C. Sim, M. Biglari-Abhari and B. C. Lovell, Face Detection on Embedded Systems, *Proceedings of the 3rd international conference on Embedded Software and Systems*, Springer Lecture Notes In Computer Science; Vol. 4523, p295-308, May 2007.
- [24] M. Tkalcic and J. F. Tasic, "Colour spaces - perceptual, historical and applicational background," in *IEEE, EUROCON*, 2003.
- [25] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Advances in active appearance models," in *International Conference on Computer Vision (ICCV'99)*, 1999, pp. 137–142.
- [26] L. Di Stefano, M. Marchionni, and S. Mattoccia, "A Fast Area-Based Stereo Matching Algorithm", *Image and Vision Computing*, pp.983 – 1005, 2004.
- [27] J. Xiao , S. Baker , I. Matthews, and T. Kanade, "Real-Time Combined 2D+3D Active Appearance Models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '04)*, pp. 535–542, 2004.
- [28] C. Hu, J. Xiao, I. Matthews, S. Baker, J. Cohn, and T. Kanade, "Fitting a single active appearance model simultaneously to multiple images," in *Proc. of the British Machine Vision Conference*, Sep. 2004.
- [29] S.C. Koterba, S. Baker, I. Matthews, C. Hu, J. Xiao, J. Cohn, and T. Kanade, "Multi-View AAM Fitting and Camera Calibration," in *Proc. International Conference on Computer Vision*, October, 2005, pp. 511 - 518.
- [30] J. Sung and D. Kim, "STAAM: Fitting a 2D+3D AAM to Stereo Images," in *Image Processing, 2006 IEEE International Conference on* , pp. 2781–2784, 2006.
- [31] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 187–194, 1999.
- [32] M.C. Ionita and P. Corcoran, "Benefits of Using Decorrelated Color Information for Face Segmentation/Tracking," *Advances in Optical Technologies*, vol. 2008, Article ID 583687, 8 pages, 2008. doi:10.1155/2008/583687.
- [33] F. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 6, pp. 567–585, June 1989.
- [34] G. Donato and S. Belongie, "Approximate thin plate spline mappings," in *ECCV (3)*, 2002, pp. 21–31.
- [35] R. K. Beatson and W. A. Light, "Fast evaluation of radial basis functions: methods for two-dimensional polyharmonic splines." *IMA Journal of Numerical Analysis*, vol. 17, no. 3, pp. 343–372, 1997.
- [36] A. Zandifar, S.-N. Lim, R. Duraiswami, N. A. Gumerov, and L. S. Davis, "Multi-level fast multipole method for thin plate spline evaluation." in *ICIP*, 2004, pp. 1683–1686.
- [37] I. Andorko and P. Corcoran, "FPGA Based Stereo Imaging System with Applications in Computer Gaming", at *International IEEE Consumer Electronics Society's Games Innovations Conference 2009 (ICE-GIC 09)*, London, UK.
- [38] M. C. Ionita, P. Corcoran, and V. Buzuloiu, "On color texture normalization for active appearance models," *IEEE Transactions on Image Processing*, vol. 18, issue 6, pp. 1372 – 1378, June 2009.



Mircea C. Ionita received the B.Sc. M.Sc.degrees in Computer Science & Electrical Engineering from University "Politehnica" of Bucharest, Romania, in 2003 and 2004 respectively. His Ph.D. degree in Image Processing & Computer Vision was awarded by NUI Galway in June 2009. His research interests include color image processing, statistical face modeling, and face recognition and tracking.



Istvan Andorko received the B. Eng. Degree in Electronic Engineering from the "Transilvania" University of Brasov, Romania in 2008. He is currently pursuing a Ph.D. degree in Image Processing & Computer Vision at NUI, Galway. His research interests include image and signal processing, VLSI and embedded systems..



Peter Corcoran received the BAI (Electronic Engineering) and BA (Math's) degrees from Trinity College Dublin in 1984. He continued his studies at TCD and was awarded a Ph.D. for research work in the theory of Dielectric Liquids. He is currently Vice-Dean of research in the Collge of Engineering & Informatics, National University of Ireland Galway. His research interests include embedded systems, home networking, digital imaging and wireless networking technologies.