# RANGE FLOW FROM STEREO-TEMPORAL MATCHING: APPLICATION TO SKINNING

Jean-Christophe Nebel and Alexander Sibiryakov
University of Glasgow, Computing Science Department
17 Lilybank Gardens, G12 8QQ Glasgow
United Kingdom

## ABSTRACT

Recently the 3D-MATIC Research Laboratory has developed techniques for the generation of 25 3D models per second of captured data. Our aim is to use these series of 3D models to study deformations of the human body. However since the 3D models have different topologies, they cannot be used directly for analyses of non-rigid motions. Therefore the generation of range flows is a prerequisite to further studies. Since we have acquired a lot of experience in correlating pixels in stereo pairs of images, we have naturally investigated the use of similar techniques to track pixels associated to range data between successive images. We present a method that allows the generation of range flows using stereo and temporal matching. We then demonstrate the efficiency of our techniques by applying the generated range flows to a study of the deformation of human skin around joints in order to perform the skinning of the region of interest.

## KEY WORDS
Range flow, Matching, Skinning, Non-rigid Motion

## 1. INTRODUCTION

For more than a decade, people of the 3D-MATIC Research Laboratory have developed techniques for the generation of 3D models of humans from stereo pairs of images. Recently we have acquired digital video cameras, which allow us to generate 25 3D models per second of captured data [1]. Our aim is to use these series of 3D models to study deformations of the human body. However since the 3D models have different topologies, they cannot be used directly for the analyses of non-rigid motion. Therefore the generation of range flows is a prerequisite to further studies.

Although the generation of optical flows has been an active domain of research for decades, the subject of range flow generation is quite recent. Moreover most of the research has been focused on rigid motions [2], [3] and motion of several objects [4]. Our work is closely related to the research done by Tsap et al., who were interested in generating range flows for non-rigid motion analysis. In particular they studied human skin deformations. They started first by investigating the use of active contours to find displacements of feature points between range maps [5]. More recently they offered a more efficient approach based on the deformation of a surface finite element (FEM) model incorporating material properties [6]. The main limitation of their original method is that it requires a priori knowledge of the soft tissue deformation: a FEM model and material properties. We should mention as well the work of Vedula et al. [7], where their starting point is not a set of range maps but a set of optical flows from 15 different cameras.

Since we have acquired a lot of experience in correlating pixels in stereo pairs of images, we have naturally investigated the use of similar techniques to track pixels associated to range data between successive images. In this paper, we present a method that allows the generation of range data and optical flows using the same matching algorithm. Using these range data and optical flows, range flows can then be generated. We then demonstrate the efficiency of our technique by applying the generated range flow to a study of the deformation of human skin around the elbow, which allows us to perform the skinning of the region of interest.

## 2. STEREO AND TEMPORAL MATCHING

The technology we use to generate range data is based on stereo-pair images collected by the camera pairs, which are then processed using photogrammetric techniques [8]. The process of finding correspondences for each pixel from a stereo-pair of images is termed stereo matching. The matching algorithm used was developed by Jin, and an earlier version is reported in [9]. Since human skin is relatively featureless at the pixel spacings and digital precisions of our current video cameras (640x480 pixels), many mismatches occur. A way of overcoming this problem is to project a random speckle pattern onto the subject. Using strobe projectors, we managed to generate series of photorealistic 3D models, at a frequency of 25 Hz by capturing the colour texture of the subject between two flashes of the strobe [1]. Another way of adding features on the subjects is to ask the subjects to wear close to the body speckle garments. Obviously that does not allow anymore the generation of models with recognizable colour textures, however that ensures sharp features, which do not depend on the projector depth of

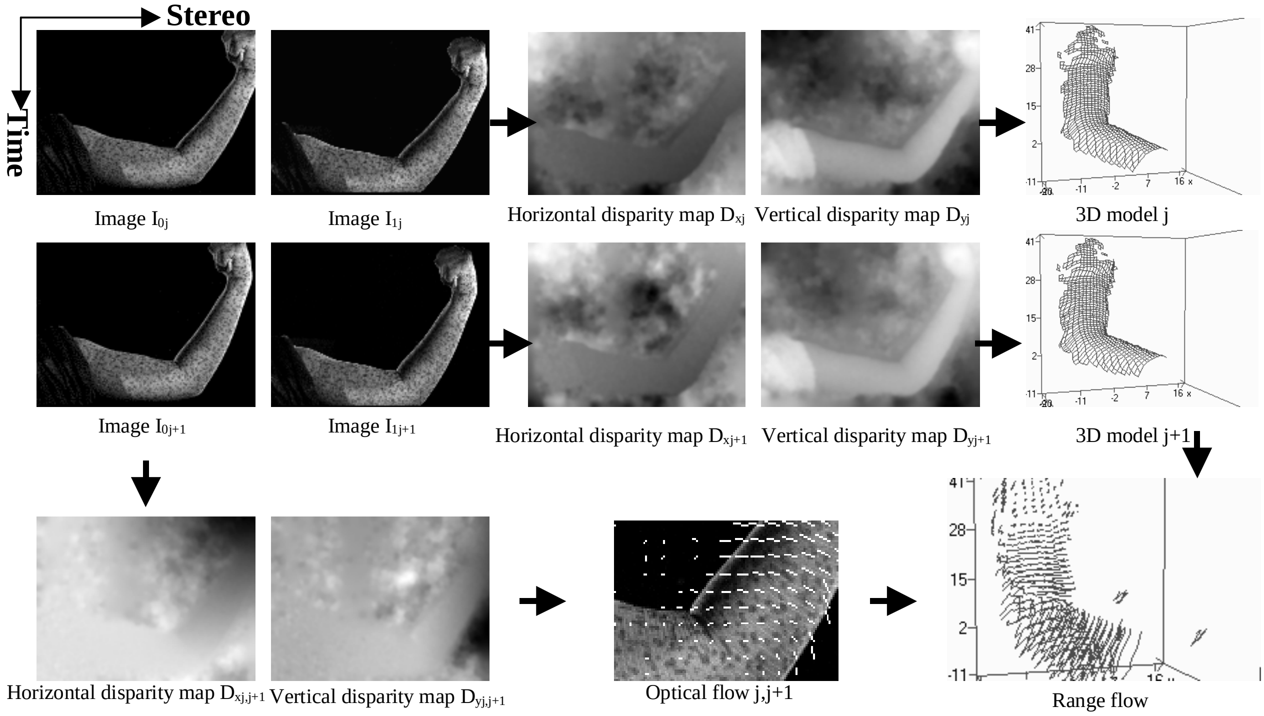



**Figure 1: Stereo and temporal matching**

field, and features which are attached to the body during the whole sequence. Therefore by tracking these features over time we could generate optical flows representing the deformation of the human skin. By combining series of range maps with their corresponding optical flows, we can then generate the range flows we need for the analysis of soft tissue deformation.

Since our stereo matcher generates horizontal and vertical disparity maps (the stereo pairs of cameras do not need to be parallel), it can be used for tracking the motion of each pixel between two successive frames generated by a given camera. Moreover each pixel has been associated with a range value during the stereo matching process; therefore the optical flows can be converted into range flows.

If *n* is the number of images captured by each stereo pair of cameras, the algorithm for generating range flows operates in two steps: First stereo matching and then temporal matching. The outer structure of the algorithm is the following:

For the *n* stereo pair of images ($I_{0j}$ and $I_{1j}$):

1. Match the images $I_{0j}$ and $I_{1j}$
2. Generate the range map $R_j$ from the disparity maps $Dx_j$ and $Dy_j$

For each temporal pair of images ($I_{0j}$,$I_{0j+1}$):

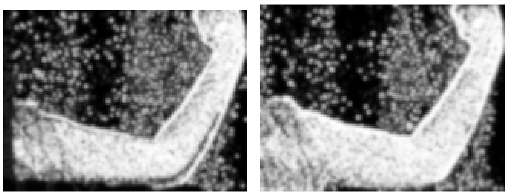
1. Match the images $I_{0j}$ and $I_{0j+1}$
2. Generate the disparity maps $Dx_{j+1}$ and $Dy_{j+1}$
3. Generate by interpolation the range value of each pixel of $I_{0j}$ at the step $j+1$
4. Iterate step 3 to calculate the range value of each pixel of $I_{00}$ at the step $j+1$

A diagram summarizing the algorithm and some results are presented on Figure 1. The range flow can be generated by temporal matching on the images $I_{0j}$. However we could generate a second range flow based on the images $I_{1j}$ to increase the accuracy of the data.

The matching algorithm itself is based on multi-resolution image correlation. We will only give a brief description of it, for more details refer to [9] or [1].

The algorithm takes as input a pair of monochrome images and outputs a pair of images specifying the horizontal and the vertical displacements of each pixel of the left image compared to the matched point in the right image. The matcher is implemented using a difference of Gaussian image pyramid: the top layer of the pyramid is 16 by 12 pixels in size for a base of 640 by 480. Starting from the top of the pyramid, the matching between the 2 pictures is computed. Then using the displacements, the right image of the next layer of the pyramid is warped in order to fit the left image. Thus if the estimated disparities from matching at the previous layer were correct, the two images would now be identical, occlusions permitting. To the extent that the estimated disparities were incorrect there will remain disparities that can be corrected at the next step of the algorithm, using information from the next higher waveband in the images. Since at each layer, the two images are supposed to match more or less, thanks to the warping step, only a neighbourhood of five by five pixels is needed for each pixel in order to find the matching pixel in the other image. Once the matching process is completed, the final displacement files combined with the calibration file of the stereo system allow the generation of a range map.

Since the matching algorithm was designed specifically to process stereo pairs of images, we need to compare the quality of the output of the temporal matching with the output of the stereo matching. This comparison is done by generating for each range map a correlation map, which gives for each pixel the correlation factor with the corresponding matched pixel at the last step of the matching. This factor takes values between 0 (black) and 1 (white), see Figure 2.



**(a) (b)**

**Figure 2: Correlation maps from (a) stereo and (b) temporal matching**

A qualitative analysis between these correlation maps does not show any significant difference between them. In many cases the correlation maps from the temporal matching look even better (with no unmatched area on the subject) than those from the stereo matching. That is due to smaller perspective distortions of the subject. For example, on Figure 2(a) the left and bottom edges are not matched because there are high perspective distortions.

The fact we need to dress our subjects with speckle clothes could appear as a big constraint of our technique for generating range flows. The alternative would be to draw a random pattern directly on the skin. This could give more accurate information about elastic properties of soft tissues. However it should be stressed we need these speckles because of the lack of features of the human skin at the level of resolution of our video cameras (640x480 pixels). Intrinsically the human skin has enough features for a good stereo matching as we demonstrated by generating range maps of the human face without any added features using a pair of 2Kx2K single shot cameras, see [10].

**(a) (b) (c) (d)**

**Figure 3. Segmentation: (a) filtering of the image of reference, (b)(c) images from the sequence, (d) result**

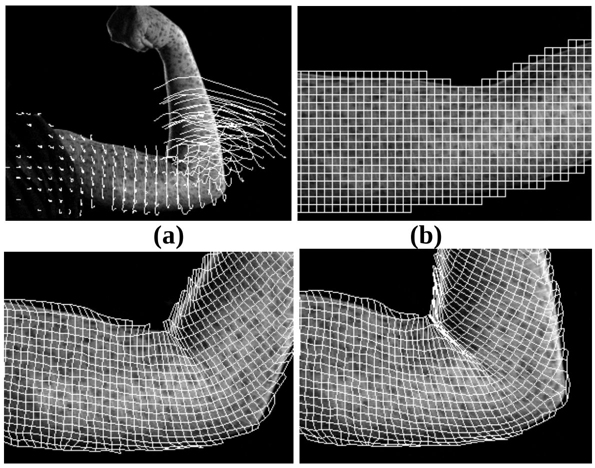
As a result of our matching method we obtain dense disparity maps. Therefore we need to extract the objects of interest from the 3D models. We segment the range flow on object/background by the following way. As we place the object in front of a dark uniform background, it appears on the disparity maps as random fields (see Figures 1,2). The segmentation is performed by applying a threshold on the correlation maps as shown on Figure 3 (a)-(c). A threshold value of 0.5 was found to be sufficient for our experiment. To filter out the random regions of high correlation in the background, we warp all binary images on the reference one using the optical flow fields. Then we take the intersection (logical AND) of all regions (Figure 3(d)). To improve the segmentation we remove small regions by morphological operations. The resulting binary mask is used to remove appropriate background vertices from the 3D models (Figure 1 - right column).

## 3. APPLICATION: 3D MODEL SKINNING

The animation of 3D characters with animation packages such as 3D Studio MAX$^{TM}$ is based on the animation of hierarchic rigid bodies defined by a skeleton. Skeletons are supporting structures for polygonal meshes that represent the outer layer or skin of characters. In order to ensure smooth deformations of the skin around articulations, displacements of vertices must depend on the motion of the different bones of the neighbourhood. The process of associating vertices with weighted bones is called skinning and is an essential step of character animation. Tools are provided by these animation packages to ease that task, however it is still a process requiring time and artistic skills.

Using the techniques previously described to generate range flows, we offer a semi-automatic method allowing an accurate skinning of scanned humans.

First we select one reference image and its corresponding 3D model from the sequence. Using the range flow we can trace each point from the reference image and obtain a deformation of any reference grid as shown on Figure 4 for the 2D case.

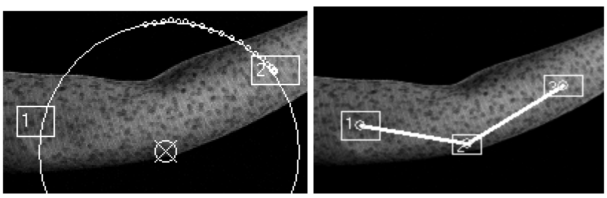


**(a) (b) (c) (d)**

**Figure 4. Using range flows for skinning: (a) point tracing, (b) the reference image with its reference grid, (c) (d) two images from the sequence with their deformed reference grids**

Our method consists in obtaining the joints of the skeleton bones and computing the weights of the points of the 3D model with respect to the bones. The manual step of the method is in the approximate selection of two regions belonging respectively to the parent bone and the child bone. Due to the direct relation between the image pixels and the vertices of the 3D model, this operation can be performed on the reference image. The rest of the process is fully automatic. Using the range flow we obtain the positions of the centre of each region in all the 3D models of the sequence. The centre and the orientation of a global coordinate system are set in the parent bone region and the positions of the centre of the child bone region are then registered in that system. We assume that the bone motion is nearly planar (so we do not consider bone bending). Therefore we can fit a plane passing through the origin of the coordinate system and all the registered positions of the child bone centres. Then we analyse the 2D-motion in that plane. First we project the positions of the child bone centres in that plane, then since the motion is circular, we can fit a circle on these points. The centre of the circle represents the 2D position of the joint and its 3D position is calculated. In Figure 5(a) the manually selected regions are shown as rectangles (the 1st rectangle represents the parent bone region and the 2nd one is the child bone region). The small circles show the positions of the centre of the child bone region in the parent coordinate system. Figure 5(a) also shows the fitted circle with its centre defining the position of the joint.

Now two 3D vectors connecting the joint and the user defined regions can be fully determined for the whole sequence (Figure 5(b)). We consider them as virtual bones because they rotate around the joint and coincide more or less with the real bones. With a wider field of view, we could have calculated by the same method the positions of the 3 joints that would have defined more precisely the positions of these virtual bones.




**(a)** **(b)**

**Figure 5. (a) Joint determination and (b) virtual bones**

The next step of the method is to assign to each vertex of the 3D model a set of weights associated to each bone. We use the following model of vertex motion:

$$\mathbf{x'} = \sum_{i=1}^{n} w_i \mathbf{R_i x_i} \qquad (1)$$

where **x'** is the deformed position of the vertex, $n$ is the number of bones, $w_i$ is the scalar weight associated to the $i$-th bone, $\mathbf{x_i}$ is the original position of the vertex in the $i$-th bone coordinate system and $\mathbf{R_i}$ is the transformation matrix of the $i$-th bone.

The 3D-rotation matrices **R** can be found from the previously calculated motions of the virtual bones. Let $\mathbf{r_0}$, and $\mathbf{r}$ be the vectors defining a virtual bone in the reference 3D model and in any other 3D model. A bone rotation can be described by an axis **p** and an angle $\alpha$.

$$\mathbf{p} = \frac{\mathbf{r}_0 \times \mathbf{r}}{|\mathbf{r}_0 \times \mathbf{r}|},\ c = \cos\alpha = \frac{\mathbf{r}_0{}^T \mathbf{r}}{|\mathbf{r}_0||\mathbf{r}|},\ s = \sin\alpha = \frac{|\mathbf{r}_0 \times \mathbf{r}|}{|\mathbf{r}_0||\mathbf{r}|}$$

and using the Rodrigues formula:

$$\mathbf{R} = \begin{bmatrix} c+(1-c)p_x^2 & (1-c)p_x p_y - sp_z & (1-c)p_x p_z + sp_y \\ (1-c)p_x p_y + sp_z & c+(1-c)p_y^2 & (1-c)p_x p_y - sp_z \\ (1-c)p_z p_x - sp_y & (1-c)p_z p_y + sp_x & c+(1-c)p_z^2 \end{bmatrix} \quad (2)$$

For simplicity we present the 2D case of the motion of a two-bone system around a joint. For this we project the 3D-vertices to the plane previously described. In this case the $\mathbf{R}_i$ are 2D-rotation matrices and equation (1) becomes:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = w_1 \begin{bmatrix} a_1 & -b_1 \\ b_1 & a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + w_2 \begin{bmatrix} a_2 & -b_2 \\ b_2 & a_2 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} \qquad (3)$$

where $w_1$ and $w_2$ can be easily found. Real motions of skin points are more complex then those expressed by the model (1). They are determined not only by the skeleton but also by muscles and soft tissue properties. Therefore the weight values obtained from (3) do not necessary satisfy to the following conditions: $\Sigma\ w_i = 1$, $0 \le w_i \le 1$. To obtain consistent values we normalise and threshold the weights:

$w_1 = w_1 / (\ w_1 + w_2)$
if $w_1 < 0$ then $w_1 = 0$
if $w_1 > 1$ then $w_1 = 1$
$w_2 = 1 - w_1$

The weights are computed for each vertex of each 3D model generated from the sequence. To obtain a smooth distribution of weights the temporal averaging of the weights of each vertex is used. Figure 6 shows the result of the weight computation. The weight distribution for the parent bone is shown in a grey palette where a white value means $w$=1 and a black one means $w$=0.

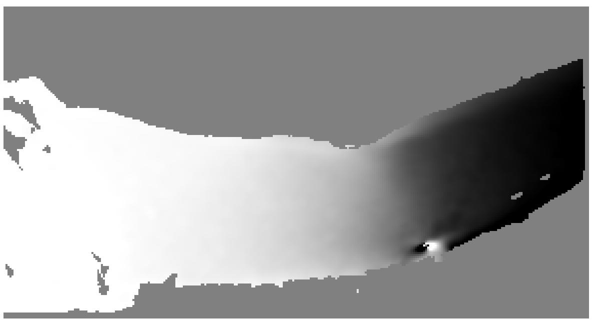


**Figure 6. Weight distribution**

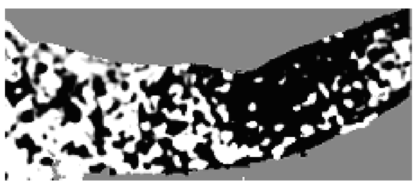
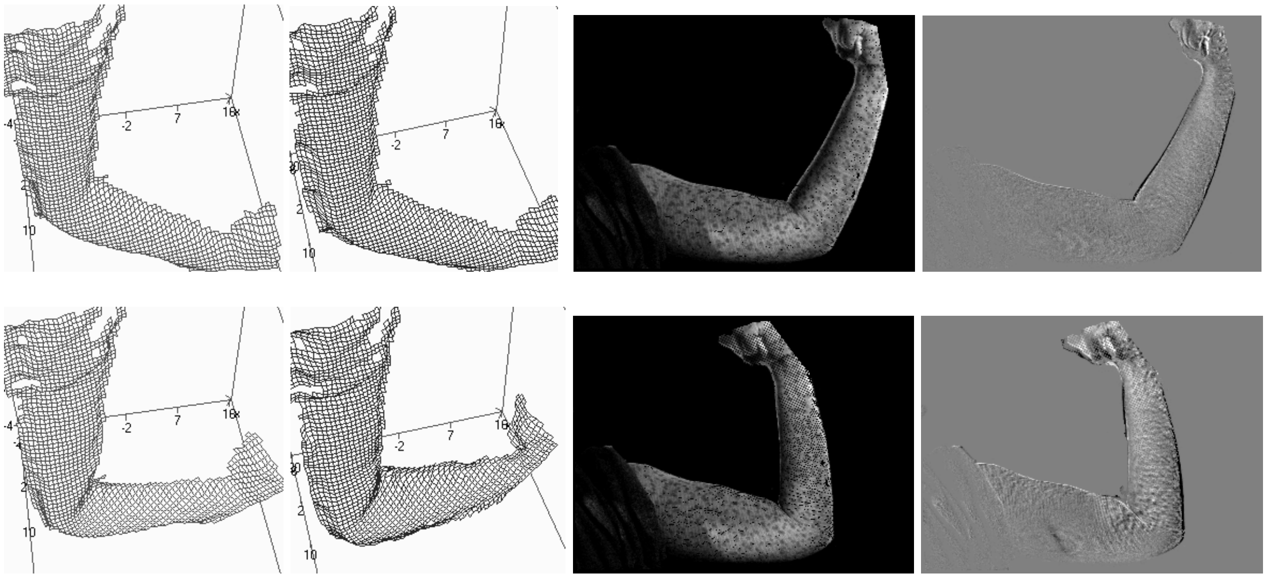


**Figure 7. Weight computation for a rigid body**



**(a) (b) (c) (d)**

**Figure 8. Two examples of skinning: (a) skinned mesh, (b) original 3D model, (c) warping of the reference image and (d) difference between the warped and the original image**

Note that in our method of weighting, the child bone must have its own rotation. Otherwise equation (2) cannot be resolved with unique values of $w_1$ and $w_2$ and as a result we obtain an almost random distribution of weights. Figure 7 shows such result when the arm is moving as a single rigid body. The same problem occurs in the joint area, it is clearly visible on Figure 6.

## 4. RESULTS

For a qualitative estimation of the validity of the computed weights, we animated the reference 3D model using techniques widely used in skeletal animation. Using the stored rotation matrices (2) of the virtual bones and the weight distribution (Figure 6), we applied the equation (1) to the reference 3D model and reconstructed the range flow. Then the reconstructed range flow was compared with the original one. This procedure is illustrated in Figure 8 (a)(b). A qualitative analysis of the real 3D model and the animated 3D model does not show any significant difference between them. The only difference is that the real one has more vertices; this is due to the motion of the object in the field of view. Therefore we can say that the animated model simulates the general motion of the skin very well and that our automated skinning process is efficient.

Another test was performed to visually estimate how our weight distribution transforms the texture. Projecting the motion of the virtual bones on the image and using the weight distribution, we applied the equation (2) to the reference image and reconstructed all the other images from the sequence by warping the reference image. Then the reconstructed images (Figure 8 (c)) were compared with the original images captured by the camera. Figure 8(d) shows the difference between them where a grey value means there is no difference and a black or a white value means a big difference. This test shows that a texture transformation by this method can also be used in animation. The transformed image does not differ much from the real one except for some illumination changes (the problem of the lighting of texture during the animation should be solved separately).

## 5. CONCLUSION

This paper presents two main results. The first one is a method of range flow generation based on stereo and temporal matching. It is expected that this method will be useful in a wide variety of applications connected with 3D motion of objects. An example of such applications would be the creation of virtual actors [11]. There are many potential further developments for our technique. For the moment the generation of range flows is based on the stereo and temporal matching which are performed independently. The method could be highly improved by using the results of the temporal matching to predict the disparity maps for the stereo matching of the next frame. This could increase the accuracy of the 3D models and reduce the computational time because the search area for the stereo matching would be heavily reduced. Also that could be useful for the segmentation of the background. However our method in its current state has already proved its efficiency by generating range flows which allowed the successful study of a non-rigid object motion: the deformation of a human arm.

The second result of the paper is the application of the generated range flows to study the deformation of human skin around joints in order to perform the skinning of regions of interest. It was shown that 2D- and 3D-analyses of a range flow could be used to obtain the weight distribution of a skin mesh. We expect in the near future that our work will be part of plug-ins written for the main commercial animation packages. The skinning task, which is still a skilled and artistic process, would become then semi-automatic and based on real data. That should contribute to cheaper and more realistic animations.

## 6. ACKNOWLEDGEMENT

We gratefully acknowledge the European Union's Framework 5 IST programme in funding this work. This work was also supported by the SHEFC project "Michelangelo".

## REFERENCES

[1] J.-C. Nebel, F. J. Rodriguez-Miguel, W. P. Cockshott, Stroboscopic stereo rangefinder, *3DIM2001*, Québec City, Canada, 2001

[2] Rigid body motion from range image sequences, *CVGIP*, *53*(1), 1-13, 1991

[3] B. Sabata and J. Aggarval, Estimation of motion from a pair of range images: A review, *CVGIP*, *54*(3), 309-324, 1991

[4] H. Spies, B. Jahne, and J.L. Barron. Dense Range Flow from Depth and Intensity Data. *International Conference on Pattern Recognition*, 131-134, 2000

[5] L. Tsap and D. Goldgof and S. Sarkar and P. Powers, A vision based technique for objective assessment of burn scars, *IEEE Trans. on Med. Imag.*, *17*(4), 620-633, 1998

[6] L. V. Tsap and D. B. Goldgof and S. Sarkar, Fusion of Physically-Based Registration and Deformation Modeling for Nonrigid Motion Analysis, *IEEE Transactions on image processing*, *10*(11), 1659-1669, 2001

[7] S. Vedula, S. Baker, P. Rander, R Collins, and T. Kanade. Three dimensional scene flow. *Proc. 7th IEEE International Conference on Computer Vision*, Kerkyra, Greece, September 1999.

[8] J. P. Siebert and S. J. Marshall, Human body 3D imaging by speckle texture projection photogrammetry, *Sensor Review*, *20*(3), 218-226, 2000

[9] J. Zhengping, On the multiscale iconic representation for low-level computer vision systems, *PhD thesis, the Turing Institute and the University of Strathclyde*, UK, 1988

[10] 3D-MATIC Research Laboratory, University of Glasgow, http://faraday.dcs.gla.ac.uk/3dmodels.htm

[11] J.-C. Nebel, Generation of True 3D Films, *LNCS 2197*, 10-19, 2001