

M. Tyshchenko

3D Reconstruction of Human Face Based on Single or Several Images

Предложен новый подход к трехмерной реконструкции человеческого лица по одному или нескольким изображениям. Подход основан на генеративной модели человеческого лица. Также предложен новый способ построения генеративной модели по набору нетекстурированных поверхностей.

A new approach to a human face 3D reconstruction by single or several images is suggested. The approach is based on a morphable model of a human face. A new way to design the morphable model by a set of untextured 3D shapes is suggested as well.

Запропоновано новий підхід до тривимірної реконструкції людського обличчя за одним або кількома зображеннями. Підхід базується на генеративній моделі людського обличчя. Також запропоновано новий спосіб побудови генеративної моделі за набором нетекстурованих поверхонь.

Introduction

A three-dimensional (3D) reconstruction of a scene based on its photographs forms a vast stratum in contemporary computer vision. State of the art provides no common technology capable of arbitrary object's automatic 3D reconstruction by a set of photographs taken from unknown viewpoints. A sparse reconstruction of rigid scenes is well studied so far. A software reconstructing camera trajectories and 3D model of a scene as a point cloud is available [1], but the subsequent triangulation of such models cannot be accurately done without human interaction [2]. Moreover, it is not always possible to achieve automatic reconstruction even significantly reducing a class of objects to be processed [3].

A 3D reconstruction of a human face occupies a separate, rather vast, niche in the indicated area. The research on this topic can be divided into two large classes: geometry oriented and morphable model oriented.

Geometry oriented approaches mostly rely on establishing the correspondence between points in different photographs of an object; surface reconstruction is then performed by determining such location of points in 3D space and such camera parameters which fit in a best way to the point correspondence. Such approaches require either substantial organizational development: special camera positioning and tuning, special illumination conditions, or significant amount of manual work:

indication of point correspondence, segmentation of scene objects etc.

Morphable model based techniques utilize the analysis-through-synthesis approach. A morphable model of a human face can be depicted as a box with tunable handles. Depending on positions of those handles the box generates one or another picture of one or another human face. The problem is to set handles of those box to such positions that it would generate exactly the input image, or at least the closest possible. A 3D shape of a face is uniquely determined by positions of those handles or, speaking more precisely, by parameters of the morphable model. An ability to reliably reconstruct a 3D shape of a human face even by a single photograph is a considerable advantage of such approaches.

1. Prior works

V. Blanz and T. Vetter's research [3] on design and use of a morphable model of human face is one of the most well known works in the area. They have managed to design an accurate morphable model and implement an efficient scheme of particular individual's face 3D reconstruction based on one or several images. Their approach is based on a conjecture that a set of all possible 3D shapes of a human face is convex in a sense. Loosely speaking, by averaging two shapes of different faces one gets a shape of some human face as well. Under this assumption, one can design a basis out of 3D shapes of several typical faces, and approximate a shape of an arbitrary face by convex combination of those basis shapes. Then a problem of human face 3D reconstruction based on its image

Keywords: 3D reconstruction, morphable model, human face, motion field, labeling problem.

consists in search for such a convex combination of basis shapes and such camera and illumination parameters, under which a generated image is least different from an input. A problem statement of such sort is a classical instance of the analysis through a synthesis approach.

Nevertheless, a method suggested by V. Blanz and T. Vetter contains a number of considerable drawbacks. The first one consists in the fact that for construction of a morphable model and for its further service in 3D reconstruction technology, the basis models have to be textured. The paper does not contain any reference for the case of untextured models. Meanwhile, bases of untextured models are much more available as compared to textured ones. Thus, it is reasonable to design a 3D reconstruction technology which makes use of just untextured models.

One more remarkable shortcoming of the technology is that for an accurate surface reconstruction one needs to specify initial approximation of orientation and illumination rather precisely. This drawback is vital for technology usage in fully automatic systems where any human interaction is impossible in principle.

A number of research efforts, carried out mainly by commercial companies, addressed to overcome the aforementioned drawbacks of V. Blanz and T. Vetter's technology.

A method, suggested in [4], is based on the detection of silhouettes on images of a face, and the search for such a convex combination of basis shapes which fits best to those silhouettes. The invariance to illumination is an advantage of this approach. But it is achieved by loss of considerable amount of information, so that an accurate surface reconstruction becomes possible only with very huge amount of images. Not to speak about 3D reconstruction of a face by a single image.

A method [5] is suggested for the reconstruction of a facial shape by series of slightly different photographs. It is founded on feature point tracking, based on which a camera position for each image is reconstructed and a rough approximation of 3D shape is built. Further, images are pairwise rectified with respect to camera positions in order to use a stereo reconstruction algorithm to refine a

model. The fact that the method works only with those series of images which can be arranged in such a way that each subsequent picture does not differ much from the previous one is a substantial drawback of the method, as well as its inapplicability in the case of a single input image.

A method [6] is designed for 3D reconstruction of a human face by two slightly different photographs and two video sequences. The method performs a search for such a convex combination of basis shapes, which fits best to location of some feature points in a pair of input images. Video sequences are used for accurate model texturing. The facts that the method works only for a pair of slightly different images, is not applicable in the case of single input image, and requires manual indication of large number of feature points are considerable shortcomings of the technique.

In the present paper we suggest a method for 3D reconstruction of a human face which does not have the aforementioned drawbacks.

2. Morphable model of a human face

A morphable model of a human face can be depicted as a box with tunable handles. Depending on positions of those handles the box generates one or another picture of one or another human face. This box contains handles of three types (fig. 1). Some of them are charged with 3D shape of the surface (α), another are charged with camera position (β), the other are charged with illumination (γ). Let us admit, that by depicting a facial shape, a camera position and an illumination by handles of that box, we virtually claim that all these quantities can be described by a finite number of parameters.

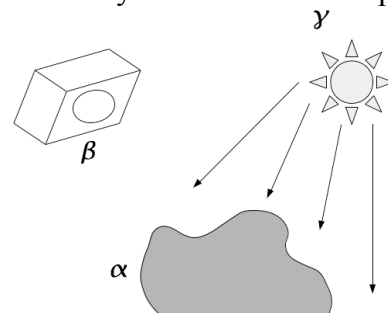


Fig. 1. Three different types of morphable model parameters

Camera position is indeed specified by finite, moreover, small number of parameters. It is not that easy in case of illumination and 3D shape. In fur-

ther subsections we will consider means to parameterize a set of 3D shapes of a human face. At present, let us consider the illumination parameterization.

2.1. Illumination model

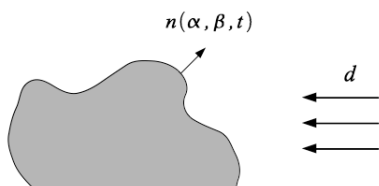


Fig. 2. Directional illumination model

Let us suppose that a surface is lit by a single directional light source, i.e. a source which shines from infinity with parallel beams of light, and is specified by a single vector d ; direction of the vector indicates light direction, and its length defines light intensity. Brightness of a surface in some point t is defined as a scalar product of normal to the surface in this point and vector d :

$$L(\alpha, \beta, d, t) = \sum_{i=1}^3 d_i \cdot n_i(\alpha, \beta, t).$$

Let us suppose that there exists so called ambient light as well. It is just a constant added to the brightness of all surface points:

$$L(\alpha, \beta, d, c, t) = \sum_{i=1}^3 d_i \cdot n_i(\alpha, \beta, t) + c. \quad (1)$$

Let us take a gander what happens if we have two directional light sources rather than one. Then the brightness of some surface point (not taking into account ambient light) equals to:

$$\sum_{i=1}^3 d_i^1 \cdot n_i(\alpha, \beta, t) + \sum_{i=1}^3 d_i^2 \cdot n_i(\alpha, \beta, t), \quad (2)$$

where d^1 is a vector, describing the first light source, d^2 describes the second one. The equation (2) can be obviously rewritten as

$$\sum_{i=1}^3 (d_i^1 + d_i^2) \cdot n_i(\alpha, \beta, t). \quad (3)$$

Thus, in this illumination model no matter we have a single light source, two, or thousand of them. All these sources are equal to some single light source.

Thus, we have parameterized illumination by a vector d and a scalar c :

$$\gamma = \{d, c\}. \quad (4)$$

2.2. Convex combination of faces

We assume that a set of 3D shapes of a human face is convex in a sense. Loosely speaking, by averaging two surfaces which are faces one gets a face as well.

Suppose we have a collection of surfaces representing some typical human faces $\{s_1, s_2, \dots, s_n\}$. We call them *basis shapes*. Let us assume that these basis shapes can be represented as points in some linear space. The means of doing so will be described in section 2.4. We approximate a shape of an arbitrary human face by a convex combination of the basis shapes:

$$s = \sum_{i=1}^n \alpha_i \cdot s_i, \quad \sum_{i=1}^n \alpha_i = 1. \quad (5)$$

Convex combination coefficients $\{\alpha_i; i=1, 2, \dots, n\}$ are just the aforementioned parameters of a facial shape α .

But what exactly should be meant by averaging of two surfaces? Let us clarify it on a very simple example.

2.3. Morphable model of rectangle

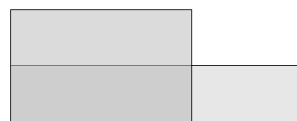


Fig. 3. Two basis rectangles

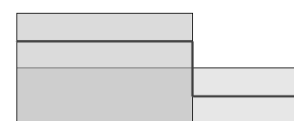


Fig. 4. Wrong averaging of basis rectangles

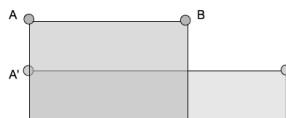


Fig. 5. Labeling of basis rectangles

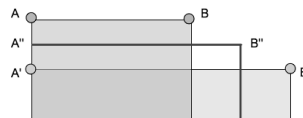


Fig. 6. Correct averaging of basis rectangles

Suppose that we would like to design a morphable model of rectangle. Let us do so in the following way: represent any rectangle as a convex combination of those, shown in fig. 3. If we consider each of introduced rectangles as a function of a single variable and average these functions, we come up with a function depicted by thick line in fig. 4, which, obviously, is not a rectangle. It has happened because we had averaged semantically different points. If we specify a proper mapping between points of the two basis rectangles, that is, map a point A of the darker rectangle to a point A' of the brighter rectangle, map a point B to

a point B' , and uniformly map all points of a segment AB to a segment $A'B'$, then we get a correct averaging (fig. 6). Indeed, a point A'' is just half way along the points A and A' , and point B'' is exactly in the middle between the points B and B' .

Thus, in order to get a rectangle by averaging two rectangles, one needs to average semantically identical, i.e. correspondent, points. This reasoning is also applicable for the case of much more complicated curves in 2D plane and surfaces in 3D space. In particular, it is applicable for the surface of a human face.

2.4. A morphable model of a human face

Let us suppose that basis shapes $\{s_1, s_2, \dots, s_n\}$ are represented as vectors containing 3D coordinates of surface points:

$$s_i = (x_1^i, y_1^i, z_1^i, x_2^i, y_2^i, z_2^i, \dots, x_{m_i}^i, y_{m_i}^i, z_{m_i}^i)^T, \quad (6)$$

where m_i – is a number of points in i -th surface.

In order to make the operation (5) formally applicable to such surfaces one has to make sure that number of points in all surfaces is the same:

$$m_i = m, \quad i = 1, 2, \dots, n.$$

In order to make the operation (5) provide a human face as a result, not an arbitrary surface, one has to guarantee that the following informally described property is also fulfilled (a formal description of the property will be given in the next subsection): components of the vectors have to be arranged in such a way that semantically identical points are placed in the same positions in their vectors. For instance, if a coordinate of a tip of the nose of the first basis shape is located in the first position in vector s_1 , then coordinates of a tip of the nose of all other basis shapes have to be located in the first positions in their vectors s_2, s_3, \dots, s_n as well (fig. 7).

Fulfilling these requirements comes easy if the correspondences between the points of different basis shapes are specified. A search for such correspondences is the major problem of fulfilling the requirements.

2.4.1. Establishing the correspondence between points of basis shapes

Let us choose one of the basis shapes, and call it the *principal shape*. If we were able to establish a correspondence between points of the principal shape and each of the other basis shapes, then we

could easily fulfill requirements, specified in p. 7. Thus we have reduced our task to a problem of establishing correspondence between points of two surfaces.

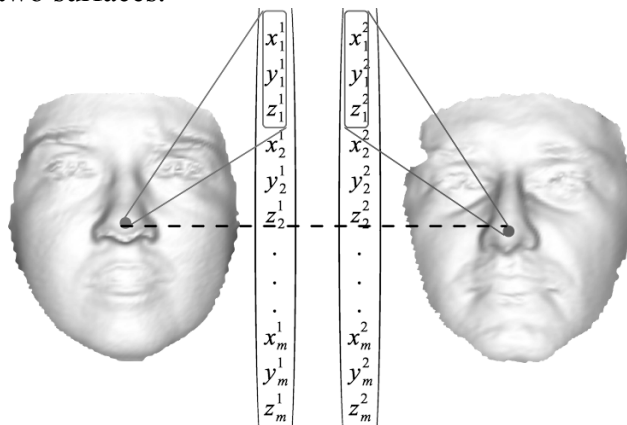


Fig. 7. Corresponding points of different 3D models have to be placed in the same positions in their vectors

On this stage we suppose that we have some fixed triangulation of surface points. Let us define a normal vector n_t as a vector of a unit length which is orthogonal to a triangle t .

We get a normal to each of the triangulation vertices by averaging normals of triangles comprising the vertex:

$$n_p = \frac{\sum_{i=1}^{k(p)} n_{t_i(p)}}{k(p)}, \quad (8)$$

where $k(p)$ is a number of triangles comprising the point p , $n_{t_i(p)}$ is a normal to i -th triangle comprising the point p . The inner part of the triangle is filled with a linear function of its vertex normals.

Normals can be visualized by ascribing to each surface point a color, which RGB components are proportional to (x, y, z) components of a normal. Frontal projections of surfaces, colored in such a way, are depicted on fig. 8. We refer to such images as to *normal maps*.

Let us remark that semantically identical points of different models have similar colors on these images. For instance, a tip of the nose is colored almost identical in the two models. In compliance with the remark, we define a difference between a point (x, y) in a normal map of the first surface and a point $(x + k_x(x, y), y + k_y(x, y))$ in a normal map of the second surface as follows:

$$\begin{aligned}
q_{x,y}(k_x(x,y), k_y(x,y)) &= \\
&= \sum_{c=1}^3 |I_1(x,y,c) - \\
&\quad - I_2(x+k_x(x,y), y+k_y(x,y), c)|,
\end{aligned} \tag{9}$$

where c is a number of RGB channel, I_1 and I_2 are normal maps of the first and the second model.



Fig. 8. Visualization of normals to a surface of a human face

Let us introduce some natural constraints on mutual alignment of the corresponding pixels in a pair of normal maps, namely forbid them to get «entangled»:

$$\begin{aligned}
&g_{i,j,i+1,j}(k_x(i,j), k_y(i,j), k_x(i+1,j), k_y(i+1,j)) = \\
&= \begin{cases} 0, & \text{if } k_x(i+1,j) \geq k_x(i,j) - 1, \\ & |k_y(i,j) - k_y(i+1,j)| < \Delta \\ \infty, & \text{otherwise,} \end{cases} \tag{10}
\end{aligned}$$

$$\begin{aligned}
&g_{i,j,i,j+1}(k_x(i,j), k_y(i,j), k_x(i,j+1), k_y(i,j+1)) = \\
&= \begin{cases} 0, & \text{if } k_y(i,j+1) \geq k_y(i,j) - 1, \\ & |k_x(i,j) - k_x(i,j+1)| < \Delta \\ \infty, & \text{otherwise.} \end{cases} \tag{11}
\end{aligned}$$

Functions (10) and (11) are equal to zero for allowed correspondence pairs, and infinity for forbidden ones. Fig. 9 visualizes these constraints: if the black pixel of the left image corresponds to the black pixel of the right image, then the grey pixel of the left image can correspond only to one of the grey pixels of the right image.

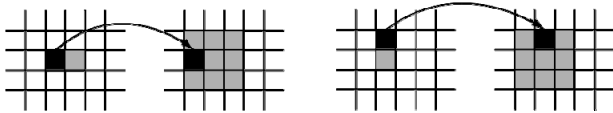


Fig. 9. Constraints on mutual location of corresponding pixels in pair of normal maps

Let us define a *penalty* for a collection of correspondences k :

$$\begin{aligned}
Q(k) &= \sum_{x=1}^X \sum_{y=1}^Y q_{x,y}(k_x(x,y), k_y(x,y)) + \\
&+ \sum_{x=1}^{X-1} \sum_{y=1}^Y g_{x,y,x+1,y}(k_x(x,y), k_y(x,y), \\
&\quad k_x(x+1,y), k_y(x+1,y)) + \\
&+ \sum_{x=1}^X \sum_{y=1}^{Y-1} g_{x,y,x,y+1}(k_x(x,y), k_y(x,y), \\
&\quad k_x(x,y+1), k_y(x,y+1)),
\end{aligned} \tag{12}$$

where k is a vector comprising a pair of numbers $(k_x(x,y), k_y(x,y))$ for each pixel of first model's normal map, X and Y are vertical and horizontal dimensions of the normal maps respectively.

A problem of optimal mapping between two normal maps consists in search for a collection of correspondences k with minimal penalty:

$$k^* = \operatorname{argmin}_k Q(k). \tag{13}$$

A problem (1) is NP-hard in general case. For the particular case, when functions $g_{i,j,i+1,j}$ and $g_{i,j,i,j+1}$ are given by (10) and (11), a polynomial algorithm is not known as well. To get an approximate solution of the problem we use a method, suggested in [7].

By solving a problem (13) we establish a correspondence between pixels of two normal maps. On the ground of these correspondences it is easy to retrieve correspondence between points of 3D models. And based on this data it is easy to fulfill the requirements indicated at p. 8. Thus by choosing different weights $\{\alpha_i : i = 1, 2, \dots, n\}$ one can generate surfaces of different human faces.

2.4.2. Expanding the possibilities of the morphable model

In order to increase a diversity of faces which our morphable model is able to generate let us split a set of facial points into four segments: eyes, nose, mouth and cheeks, as it is depicted in fig. 10. Let us introduce a set of convex combination coefficients $\{\alpha_i^S : i = 1, 2, \dots, n\}$ for each of the segments

$S \in \bar{S}$. It is obvious that the arbitrary chosen coefficients α produce gaps on borders of the segments i.e. we get for disconnected segments of a face rather than a single face. We use the following technique in order to eliminate those gaps.



Fig. 10. Face segmentation

For each segment $S \in \bar{S}$ and for each point p of the surface calculate a distance from this point to the nearest point of the segment:

$$d_s(p) = \min_{v \in S} |p - v|. \quad (14)$$

Normalize this distance over all segments and over all points of the surface:

$$\hat{d}_s(p) = \frac{d_s(p)}{\max_{S' \in \bar{S}} \max_p d_{S'}(p)}. \quad (15)$$

Calculate a coefficient of point p belonging to a segment S :

$$b_s(p) = (1 - \hat{d}_s(p))^4. \quad (16)$$

The fourth power is chosen to guarantee a sufficient attenuation speed for the function $b_s(p)$ while moving away from the segment border. Normalize this coefficient:

$$\hat{b}_s(p) = \frac{b_s(p)}{\sum_{S' \in \bar{S}} b_{S'}(p)}. \quad (17)$$

Calculate a convex combination coefficient for point p :

$$\alpha_i(p) = \sum_{S \in \bar{S}} \alpha_i^S \cdot \hat{b}_S(p). \quad (18)$$

Convex combination coefficients of the basis shapes are now different for different points of the model. Such a modification substantially increases a diversity of faces produced by the morphable model.

3. 3D reconstruction of a human face based on a photograph

In the previous section we have in depth researched the interior of the morphable model. It is now convenient for us to return to such an abstraction level where a morphable model is just a box with three types of parameters: those charged with 3D shape of the surface (α), those charged with ca-

mera position (β), and those charged with illumination (γ). Under particular values of these parameters a morphable model generates a picture of a human face with a 3D shape given by parameters α , taken under a viewpoint, given by parameters β , and illumination conditions, given by parameters γ .

This is the time to ask a question: what exactly this generated image looks like? Let us recall that our morphable model generates a non-textured surface. For certainty let us suppose that all points of a surface are colored in white. Then in the generated image we get something similar to a plaster statue photograph. Let us first describe a method of 3D reconstruction for the case when we have a photograph of a plaster statue of a human face as an input image. Afterwards we will consider a way to modify the suggested algorithm to work with photographs of real human faces.

3.1. 3D reconstruction of a plaster statue

Suppose we have an input image

$$I: T \rightarrow RGB, \quad (19)$$

where T is a set of pixels, $RGB = \{0, 1, \dots, 255\} \times \{0, 1, \dots, 255\} \times \{0, 1, \dots, 255\}$.

Suppose we have an image, generated by the morphable model with respect to parameters α , β and γ under illumination model (1):

$$L_{\alpha, \beta, \gamma}: T \rightarrow RGB. \quad (20)$$

Let us define a difference between the input and the generated image:

$$F(I, L_{\alpha, \beta, \gamma}) = \sum_{t \in T} (I(t) - L_{\alpha, \beta, \gamma}(t))^2, \quad (21)$$

where T is a subset of pixels, into which at least one of the surface points is projected.

The problem of 3D reconstruction of a human face based on its image consists in looking for such parameters α , β and γ , under which a difference between the input image and the generated one is minimal:

$$F(I, L_{\alpha, \beta, \gamma}) \rightarrow \min_{\alpha, \beta, \gamma}. \quad (22)$$

Given fixed parameters α and β a function (22) under illumination model (1) is minimized with respect to γ by least squares. The minimization with respect to α and β is performed by Nelder–Mead method [8]. To make the paper self-contained we give a brief description of the method.

Suppose a continuous function $f: R^n \rightarrow R$ is given. The Nelder–Mead algorithm maintains a simplex $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{n+1}\}$ in R^n . The algorithm requires a termination threshold ε as well as parameters α , γ , δ and σ which are explained later on. The algorithm performs the following sequence of operations:

1. *Order* all simplex vertices according to the values of function f : $f(\mathbf{x}_1) \leq f(\mathbf{x}_2) \leq \dots \leq f(\mathbf{x}_{n+1})$.

2. Calculate \mathbf{x}_0 , the center of gravity of all simplex points, except \mathbf{x}_{n+1} .

3. *Reflection*. Compute a reflected point:

$$\mathbf{x}_r = \mathbf{x}_0 + \alpha(\mathbf{x}_0 - \mathbf{x}_{n+1}).$$

If: $f(\mathbf{x}_1) \leq f(\mathbf{x}_r) \leq f(\mathbf{x}_n)$,

then substitute a point \mathbf{x}_{n+1} by \mathbf{x}_r , and go to step 1.

4. *Expansion*. If the reflection point is the best so far:

$f(\mathbf{x}_r) \leq f(\mathbf{x}_1)$ then compute an expansion point:

$$\mathbf{x}_e = \mathbf{x}_0 + \gamma(\mathbf{x}_0 - \mathbf{x}_{n+1}).$$

If the expansion point is better than the reflection point:

$$f(\mathbf{x}_e) \leq f(\mathbf{x}_r)$$

then substitute the worst point \mathbf{x}_{n+1} by expansion point \mathbf{x}_e and go to step 1. Otherwise, substitute the point \mathbf{x}_{n+1} with point \mathbf{x}_e and go to step 1.

5. *Contraction*. On this stage it is certain that:

$$f(\mathbf{x}_r) \geq f(\mathbf{x}_n).$$

Compute a contraction point:

$$\mathbf{x}_c = \mathbf{x}_{n+1} + \zeta(\mathbf{x}_0 - \mathbf{x}_{n+1}).$$

If the contraction point is better than the worst point:

$$f(\mathbf{x}_c) \leq f(\mathbf{x}_{n+1})$$

then substitute a point \mathbf{x}_{n+1} with \mathbf{x}_c , and go to step 1.

6. *Reduction*. For all point, except the best one, assign:

$$\mathbf{x}_i := \mathbf{x}_1 + \sigma(\mathbf{x}_i - \mathbf{x}_1), \quad i \in \{2, 3, \dots, n+1\}.$$

Calculate distances between the best point and all other simplex vertices: $\delta_i = \|\mathbf{x}_i - \mathbf{x}_1\|$, $i \in \{2, 3, \dots, n+1\}$.

If $\max_i \delta_i > \varepsilon$, go to step 1. Otherwise, end up with a point \mathbf{x}_1 .

Standard values for parameters α , γ , ρ and σ are the following: $\alpha = 1$, $\gamma = 2$, $\rho = 1/2$, $\sigma = 1/2$.

3.2. 3D reconstruction of a human face

In order to process photographs of real human faces, rather than images of single-colored plaster statues, we remove all non-skin segments from the image, i.e. eyebrows, eyes and lips, as it is shown in fig. 11. As semantically identical points of all basis shapes are located in the same positions in their vectors, it is sufficient to remove the indicated segments only from one of the basis shapes. This operation is then mechanically transferred to the other models.



Fig. 11. Non-skin segments removal

Fig. 12 and 13 show examples of human face 3D reconstruction made by the suggested method.

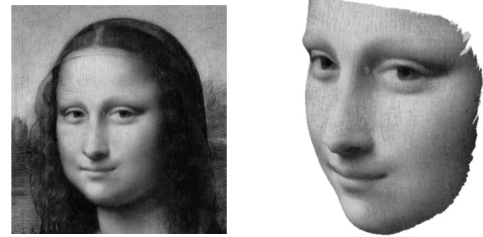


Fig. 12. An example of human face 3D reconstruction by single image Fig. 12 and 13 show examples of human face 3D reconstruction made by the suggested method.

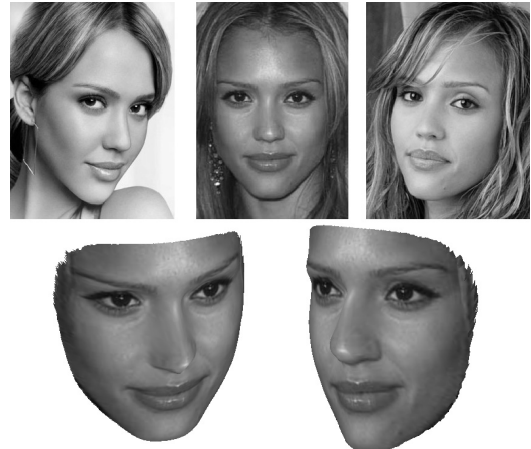


Fig. 13. An example of human face 3D reconstruction by several images

1. *2D3 LTD*, 2008. Boujou 4: The virtual interchangeable with real. – <http://www.2d3.com>
2. *Unwrap mosaics: a new representation for video editing* / A. Rav-Acha, P. Kohli, C. Rother, A. Fitzgibbon. // Proc. of SIGGRAPH 2008..
3. *Blanz V., Vetter T.* Face identification across different poses and illumination with a 3D morphable model // IEEE Trans. Pattern Analysis and Machine Intelligence. – 2003. – **25**, – P. 1063–1074.
4. *US Patent N 7212664.* Mitsubishi Electric Research Laboratories, Inc. Constructing heads from 3D models and 2D silhouettes.
5. *US Patent N 7103211.* Geometrix, Inc. Method and apparatus for generating a 3D face models from one camera.
6. *US Patent N 7212656.* Microsoft Corporation. Rapid computer modeling of faces for animation.
7. *Kolmogorov V.* Convergent Tree-Reweighted Message Passing for Energy Minimization // IEEE Transactions on Pattern Analysis and Machine Intelligence. October 2006. Washington, DC, USA. – **28**. – P. 1568–1583.
8. *Nelder J.A., Mead R.* A simplex method for function minimization // Computer J., 1965. – **7**. – P. 308–313.

© M. Tyshchenko, researcher, the International Research and Training Centre of Informational Technologies and System, Kiev, 2011