Computer Graphics Project: 3-D Facial Hair Modeling

Shivam Srivastava (CS10B051)
K. Vignesh (CS10B054)

Introduction:

Mastering the human face has been a goal in computer graphics for a long time. It is only recently that the focus has shifted to the finer details such as the facial hair. Most of the prior art regenerates the facial hair as a single mass without paying attention to individual components. This project focuses on detecting and reconstructing facial hair (eye brows, eye lashes, moustache and beard) in 3-D. The following sections describe the algorithm used, results obtained and things not achieved.

Problem Statement:

The aim of the project is to use algorithms to detect and trace hairs in the captured images and reconstruct them in 3-D using a multi-view stereo approach.

So, given an input of a face from which the facial hair needs to be detected, the output should be a 3-D facial hair model with 3-D facial hair.

Background:

Before delving into the algorithm a few terms need to be understood.

3-D reconstruction: A single camera is capable of storing data only on a 2-D surface. But, with multiple cameras the 3-D view can be reconstructed using appropriate calibration techniques. This method of reconstructing the original 3-D model using only 2-D images and calibration is called 3-D reconstruction.

Multiview Stereo: The approach used for 3-D reconstruction is more formally called Multiview Stereo. The setup involves placing cameras at different positions with respect to the subject and then calibrating them with respect to each other using various techniques. The output includes transformation matrices which will be used in our algorithm described below.

Gabor Filter: It is mainly used for filtering out edges from images. As the level at which we are capturing hair is rather small, each hair can be treated as a 1-D segment. Hence, as will be described later, this filter is used to filter out hair from an input image.

Skin Episurface: It is a term introduced in the paper we have implemented. For regions with sparse hair, the skin episurface is a close approximation of the true underlying skin surface. For regions with dense hair, the episurface is a postulated 3-D surface below the top layer of visible hair. The motivation behind using this is that it allows one to use a unified approach for hair reconstruction and not be concerned about the variations at different regions of the face. Figure 1 describes the concept.

Shrink Wrap Surface: A surface can be called shrink wrapped if, it has features which are non-static with respect to a base surface but during reconstruction were left undifferentiated from the underlying surface, leading to an exact wrapping of the surface in a way.

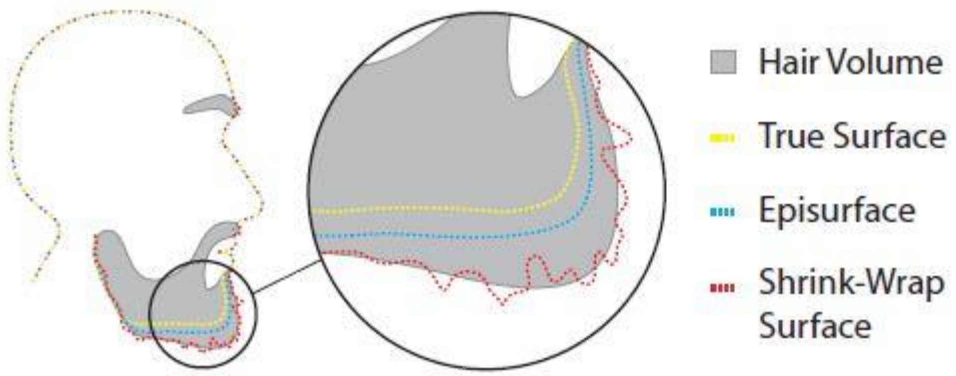


Figure 1: Episurface (Image courtesy: Disney Research Group)

Inpainting: It is the process of reconstructing lost or deteriorated parts of the image. In our case, once the hair is separated from the image, inpainting is used to reconstruct the underlying skin surface.

<Add more terms as and when new terms pop up in the remaining report>

Algorithm:

The following figure describes an overview of the entire pipeline for the 3-D reconstruction. (Change the following image to your own)

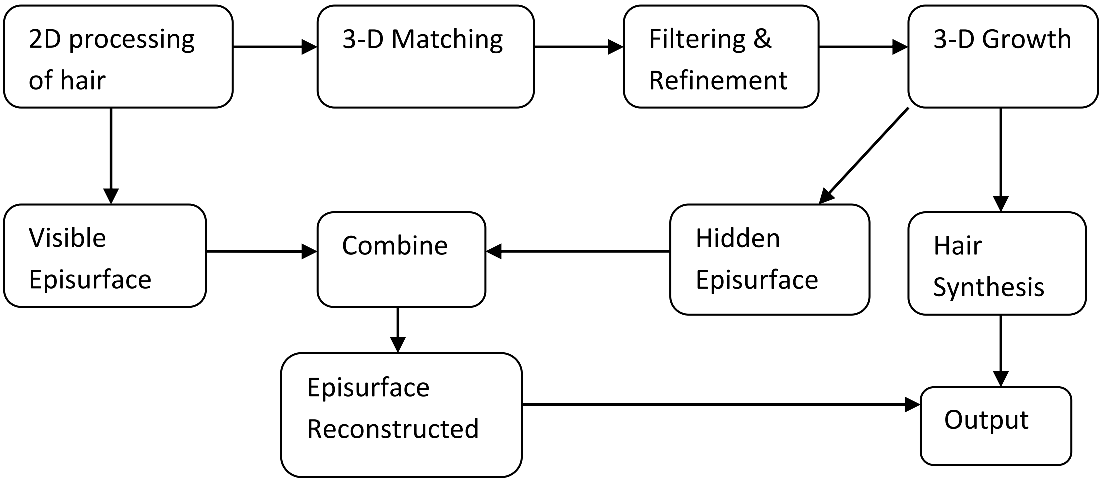


Figure 2

The approach described in [1] uses the above algorithm. The algorithm works by explicitly differentiating the skin episurface and the hair, and processing each of them in a distinct way. The pipeline is as follows:

a) Separate the hair and skin in the captured images, extract 2-D hair fibers using a growing algorithm and remove the detected hairs from the image using inpainting.
b) Reconstruct, filter, refine and grow 3-D fibers in 3-space based on the extracted 2D fibers using multi-view stereo (MVS).
c) Compute the skin episurface by combining the traditional MVS (as shown in [2]) and the estimated roots of the 3-D fibers.
d) Synthesize hairs in areas where image data indicates the presence of hair, but individual hairs are indistinct and cannot be reconstructed.

Before implementing any of the above steps, a mesh of the shrink wrapped surface of face is constructed using the ideas in [2]. The following will give a step by step description of what we have done in the project.

Work Done

Data:

The data was obtained from Disney Research Lab, Zurich. It consists of 14 images captured from different cameras whose calibration is also given.

2-D Processing:

The first step is to detect hair in the input images and trace piecewise linear segments of hair.

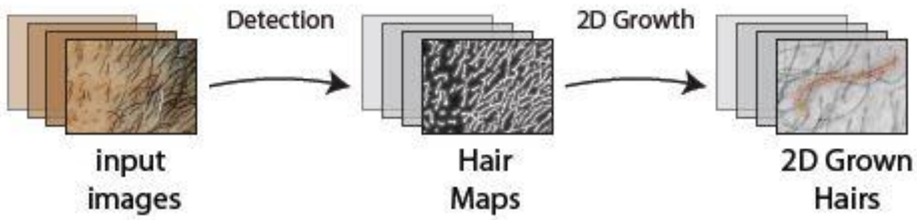


Figure 3: Detecting and growing hair in 2-D (Image: Disney Research Lab)

1) **Computing a 'Hair Map' for an image**: Neither the hair color nor does its diameter remain the same across all faces, so these methods are eliminated for detection. It has been observed in [1] that hair and skin exhibit larger contrast in saturation and value than in hue. Hence, the images are converted from RGB to an HSV space and the H and V channels are used to discriminate hair.

   The next step is to apply oriented filters which have been shown to detect local orientation of hair, well. These need two priors. First, the size of the structural element to filter (hair thickness) and second, the angle at which to filter. As the hair thickness can be estimated roughly, the only parameter to be estimated is the angle. This filter kernel is applied after the H and V channels are used.

   Different orientations of the kernel are used and the angle which produces the highest score:
   $$F(x,y) = |K_{\tilde{\theta}} * V|_{(x,y)} + |K_{\tilde{\theta}} * S|_{(x,y)}$$
   at (*x*,*y*) is stored in the orientation map O(*x*,*y*). After applying the filter, the image still has noise which needs further filtering. To suppress these, a non-maximum suppression strategy is used. Along any hair strand, only the point which has the maximum score along a direction orthogonal to *theta* is stored in a new map called the confidence map (C). The confidence map is

thresholded (real values are converted to binary values) between 0.05 and 0.07 to obtain the binary mask (M).

Finally, from the binary mask (M), a hair map H is generated using the following formula:

$$H(x,y) = 1/(1+d(x,y))$$

Here, $d(x,y)$ is the Euclidean distance at (x,y) to the closest pixel in M with value 1. So, in H, the value is 1 at (x,y) where a hair is suspected and it decays quickly when moving away from the suspected hair pixel.
Hair maps will be used for further growth and reconstruction. There are high chances that across different views (cameras), individual hair segments are not detected appropriately. For this reason, hair maps are extracted for further matching and reconstructing. A step by step construction is as follows:

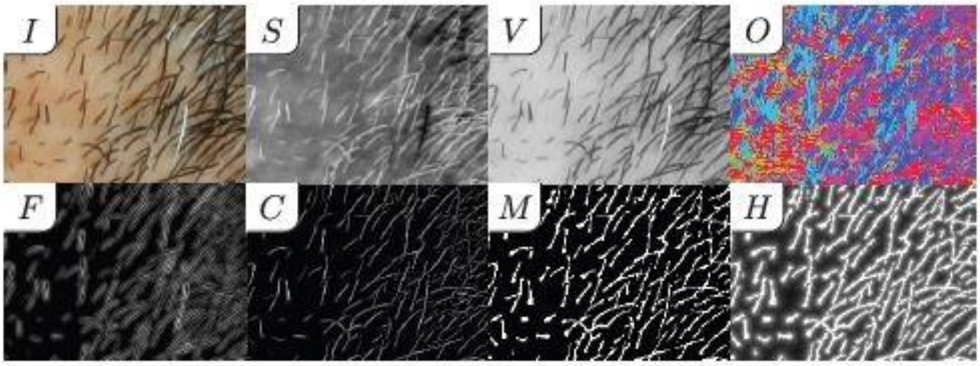


Figure 4: *Images used during reconstruction: I: input image, S: saturation channel, V: value channel, O: orientation map, F: Gabor filter response, C: confidence map, M: binary mask, H: hair map.*

(*Image: Disney Research Labs*)

2) **Growing hair in 2-D:** At the scale we are detecting hair, they are 1 dimensional segments. Hence, for further matching and reconstruction, the only neighborhood that needs to be considered is along this 1-dimensional structure.
A chain of 2-D hair segments is formed on growing in 2-D using the hair maps found above. A hair segment $s(p_s, \theta, l)$ is a linear segment of length 'l' which starts at p_s and in the direction of θ. $P(s)$ denotes the set of pixels covered by the segment *s*. Growing works as follows:
   i) For each pixel p_s for which $H(p_s) = 1$, that has not been used yet, is selected as a seed and the growing direction is determined using the orientation map *O*.

ii) A cone $f(r,\gamma)$ is defined where ‘r’ is the growth resolution and the opening angle of the cone is $2*\gamma$. Refer figure 5. The parameters used for the growth cone are 10 pixels and 60 degrees respectively.

iii) For each potential growth direction in the growth cone, a score is computed using:

$$\xi(d\theta) = \left(1 - \frac{|d\theta|}{2\gamma}\right) \psi(\mathcal{P}(s_{d\theta}))$$

Here, $d\theta$ is the angular direction relative to the axis of the cone and $P(s_{d\theta})$ denotes the set of pixels covered by the segment $s_{d\theta}$. The function ψ is defined as the following:

$$\psi(\mathcal{P}) = \frac{1}{\|\mathcal{P}\|} \sum_{p_i \in \mathcal{P}} \frac{H(p_i) - \nu}{1 - \nu}$$

Here, ν controls how tolerant the score is. Its range is [0, 1). The higher it is, the more restrictive the score becomes, but at the same time it becomes less robust also. Refer figure 5 (b) below for details.

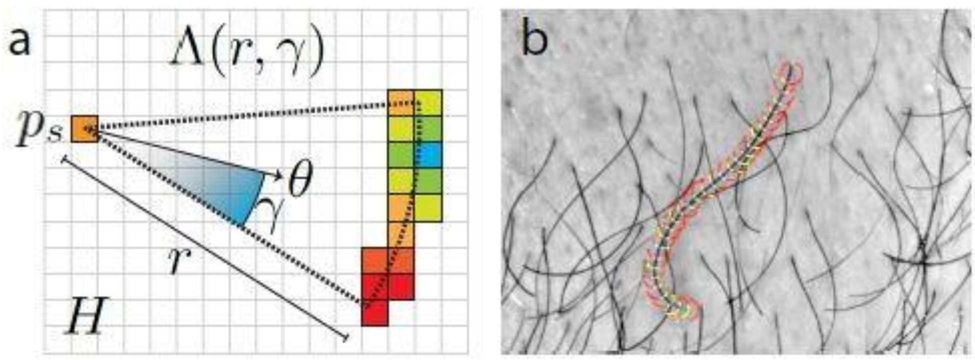


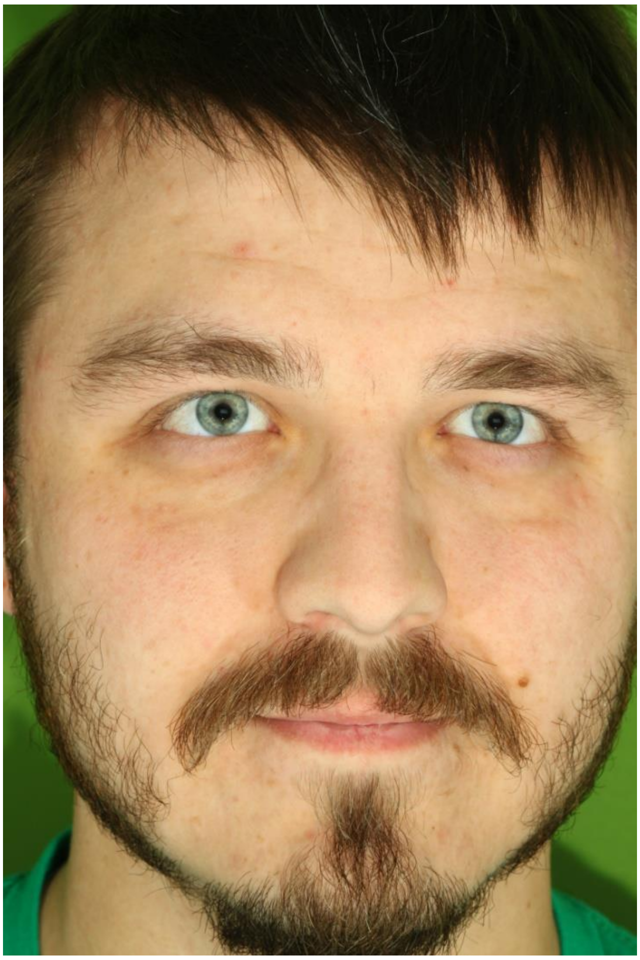
Figure 5: *2D hair growth. Hair growing in 2D makes use of the hair map H. (a) Starting point p_s and an initial estimate of the growth direction θ are given by the previous segment. The apex of a growth cone $\Lambda(r,\gamma)$ with growth resolution r and opening angle 2γ is placed at p_s and oriented along θ. For all possible target pixels a score is computed and the pixel with highest score is added to the hair. This process is repeated until the matching score drops below a threshold. (b) An example of a traced hair overlaid the input image.*

(*Image courtesy: Disney Research Labs*)

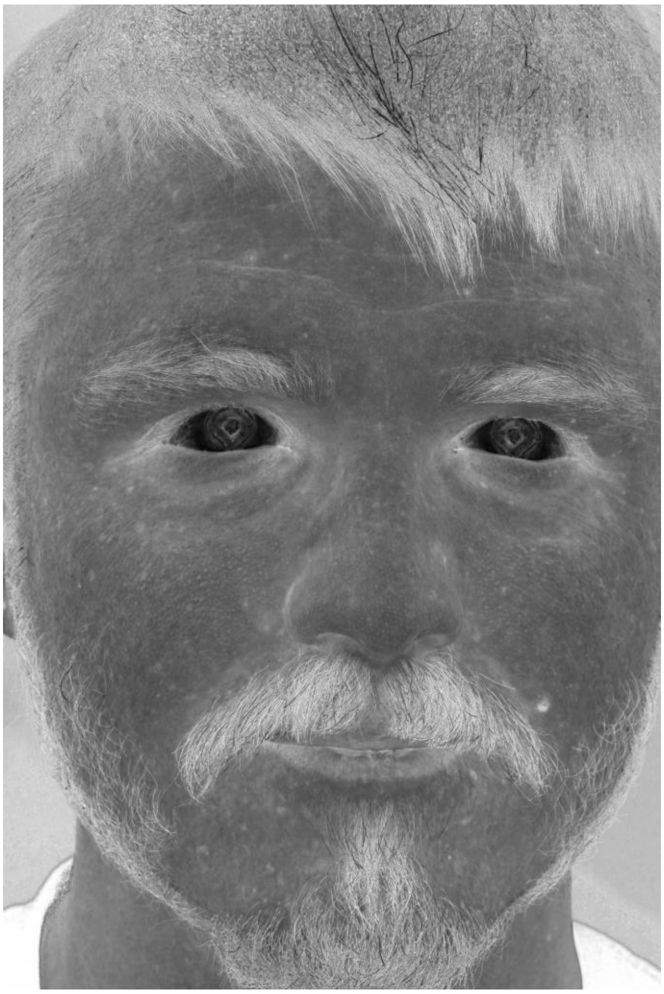
3) **Calculating disparity maps for constructing shrink wrap surface:** For this step, [2] was used. This step could not be completed. Disparity maps between images could be found and a point cloud was generated using these. The images obtained are shown in the *result* section.

Results:

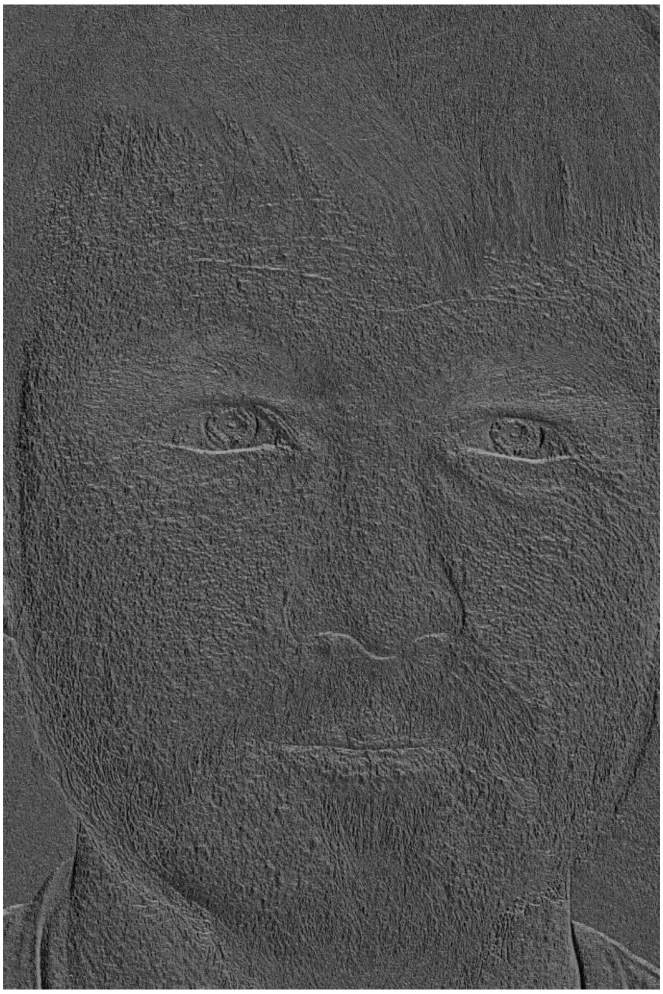
1) *Hair map generation*:



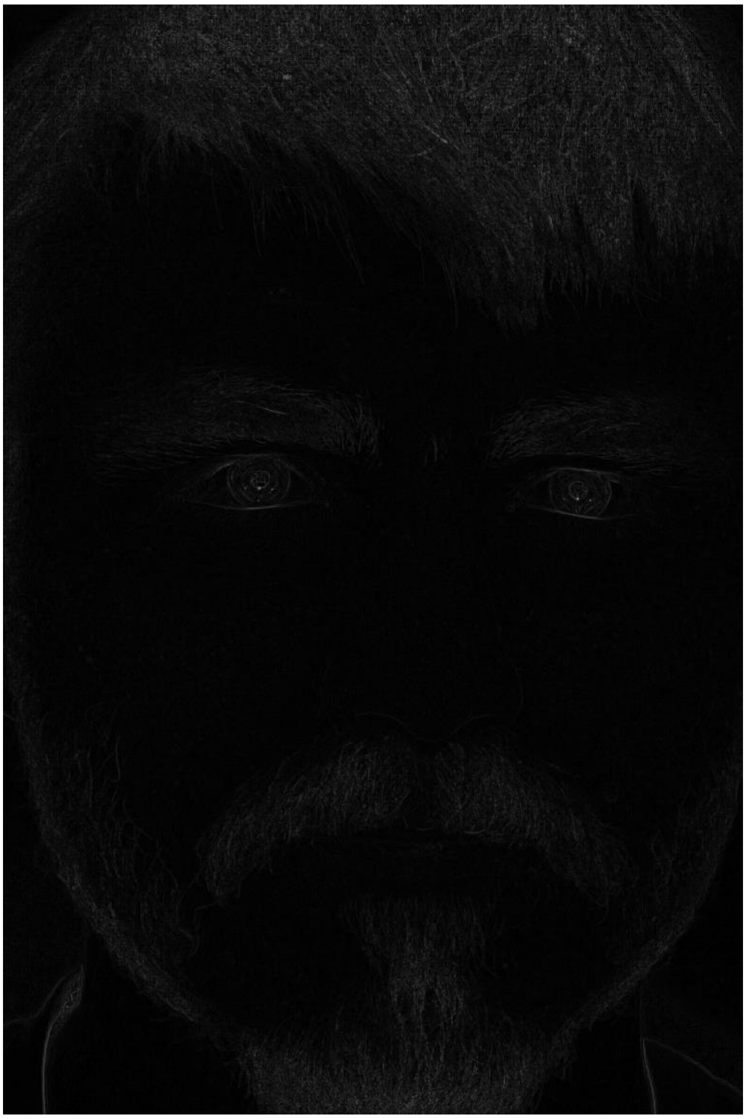
(Original image – 0)

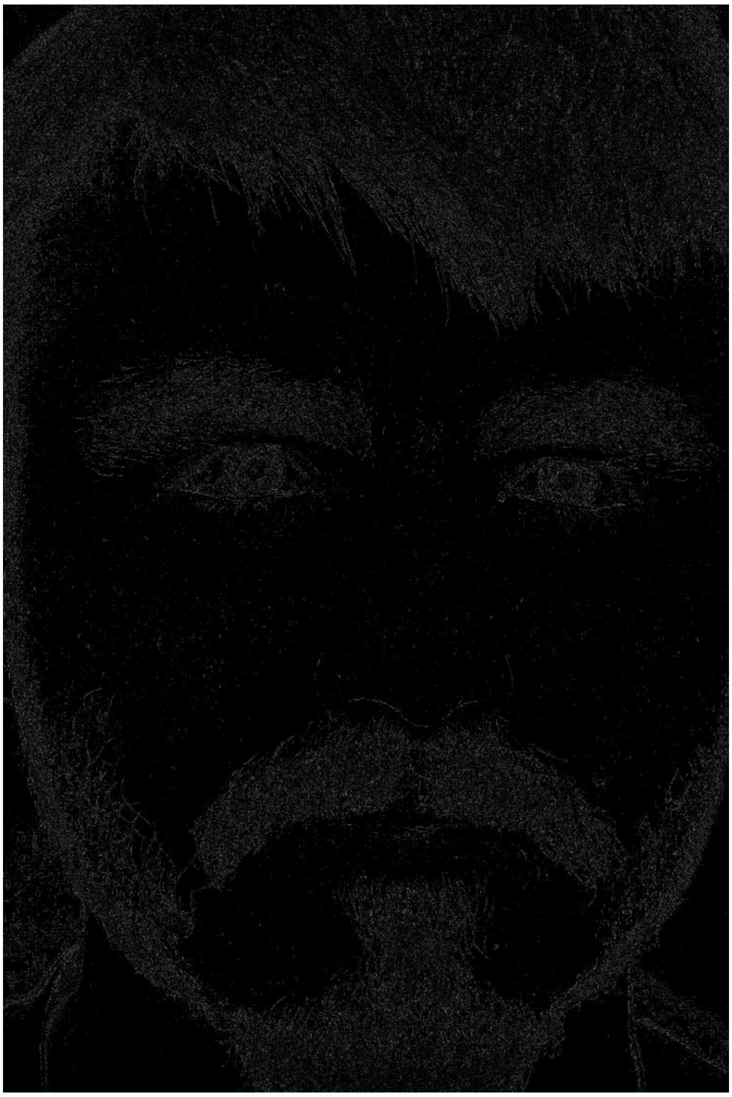

(S component of the HSV image of image 0)

(Value part of the HSV of image - 0)

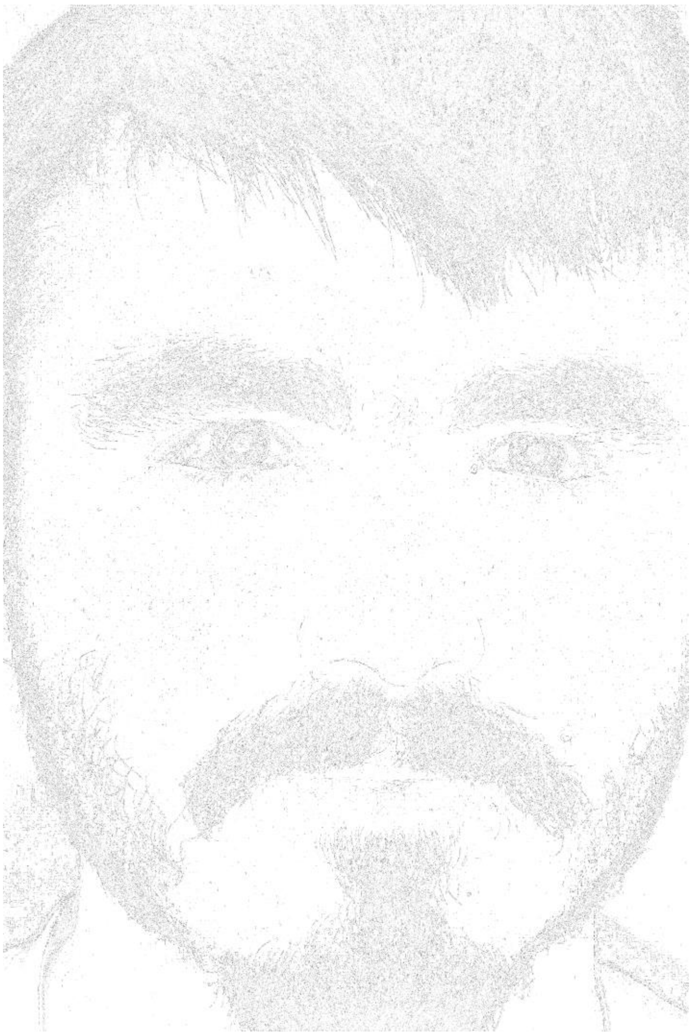


(Orientation Map image for image 0)

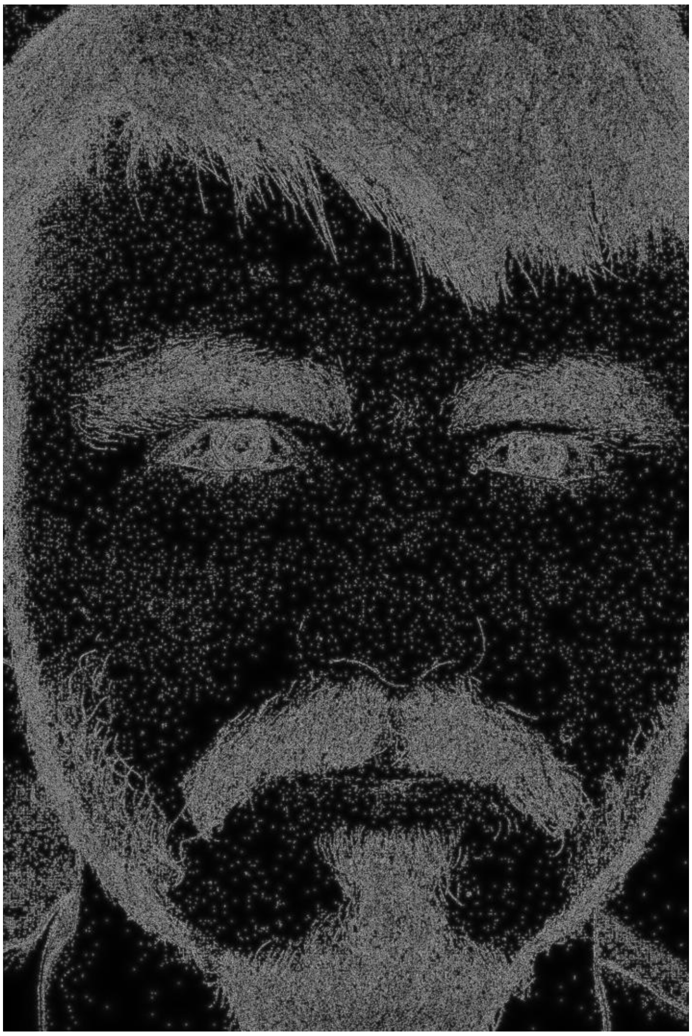

(Gabor Filter response)



(Confidence map for Image-0)

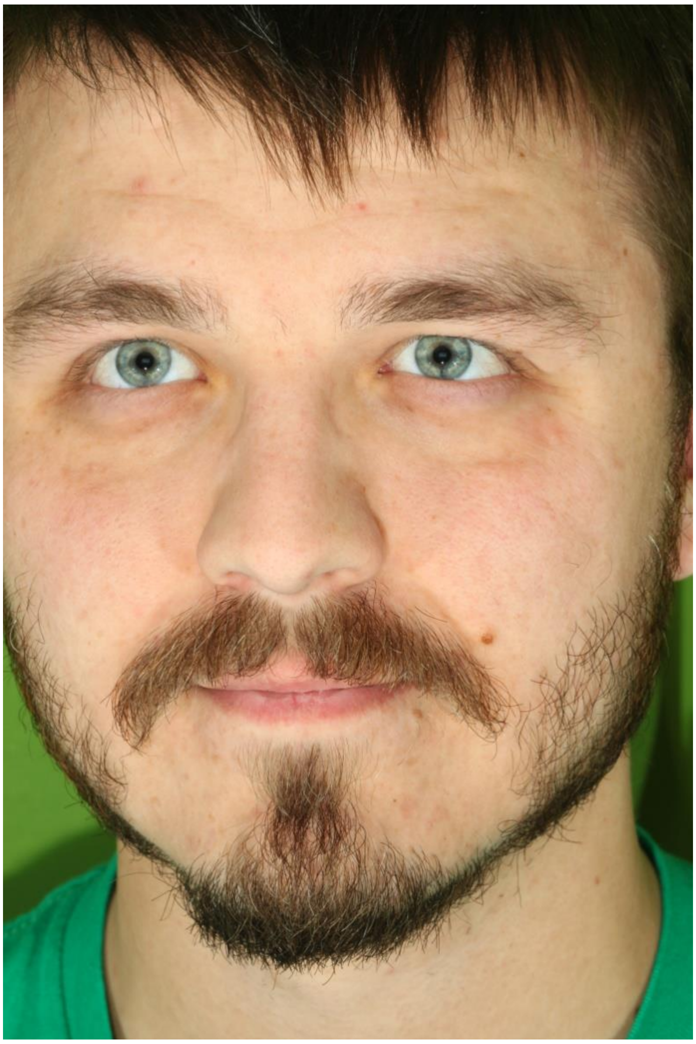


(Output of non-maximum suppression strategy)

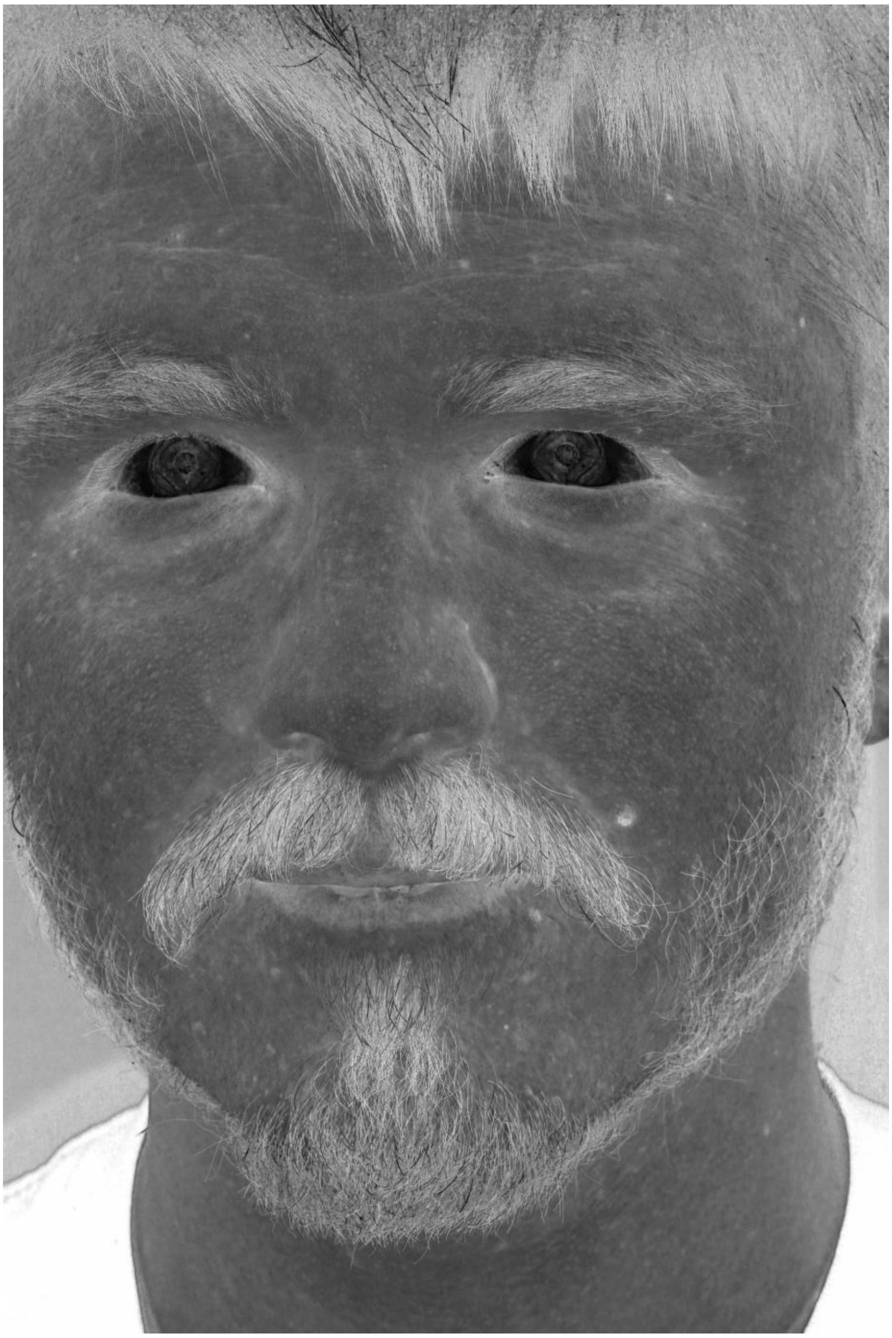


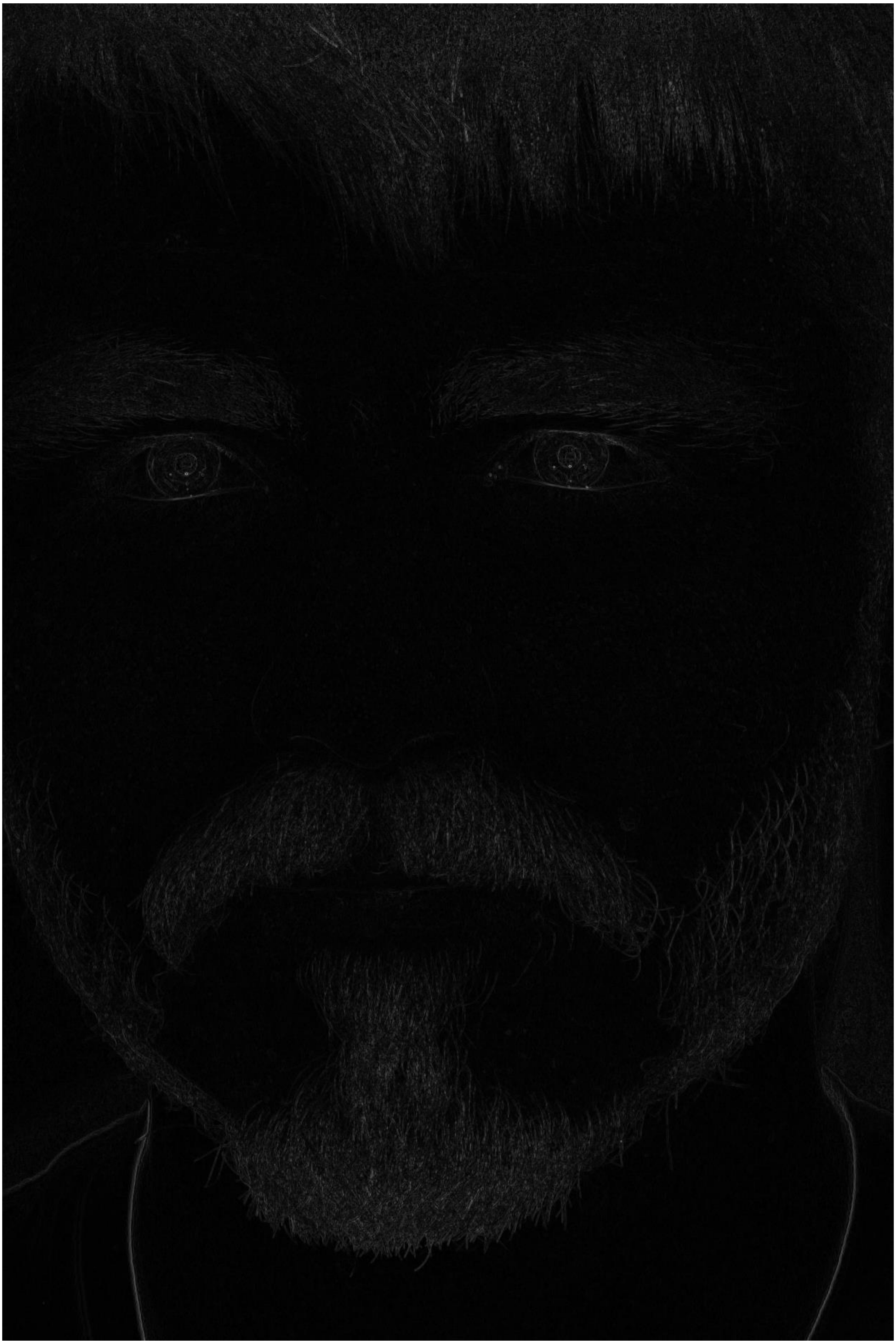
(Hair map for image 0)

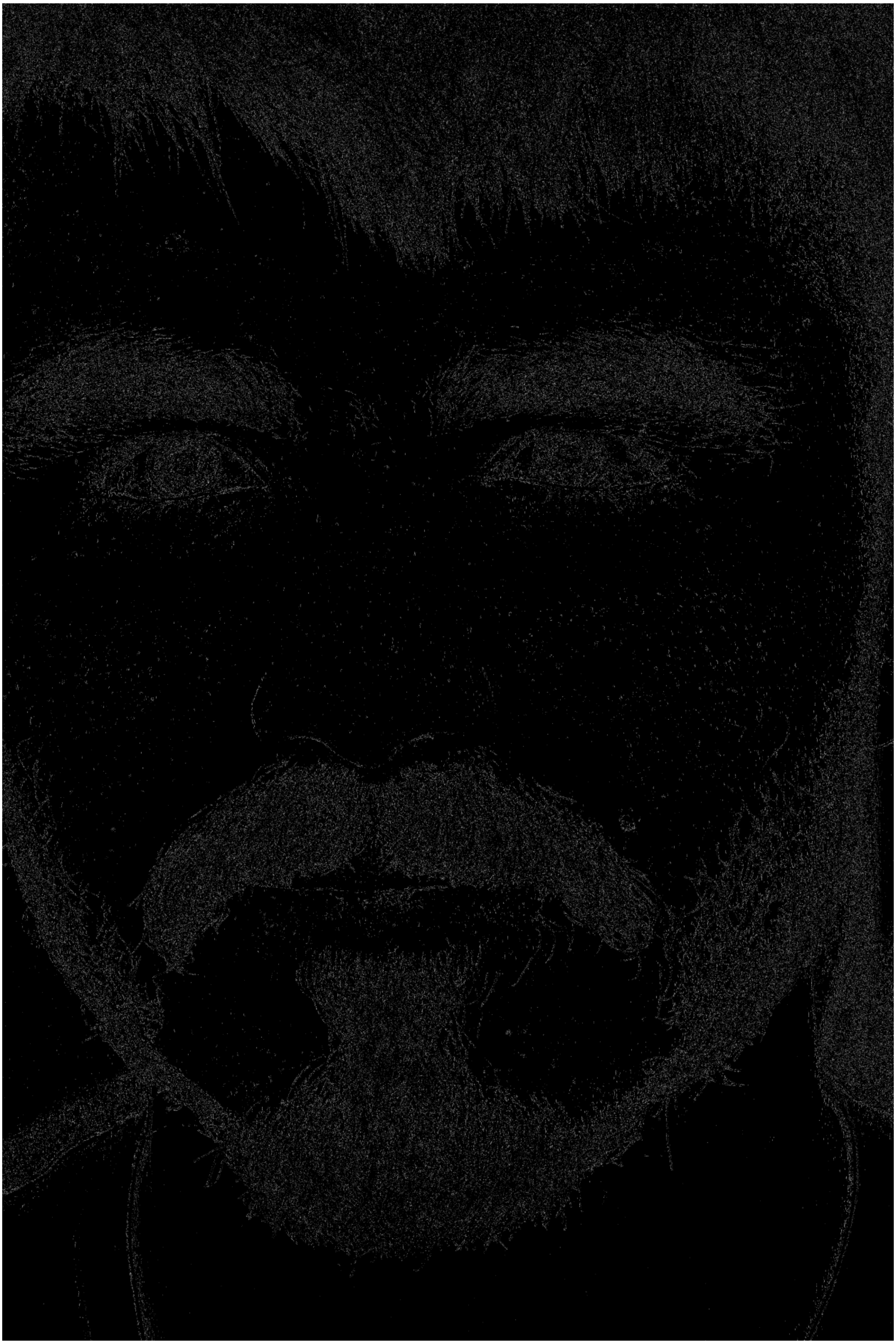
Hair map generation for Image 1
(The same order as above, hence no description for each image)

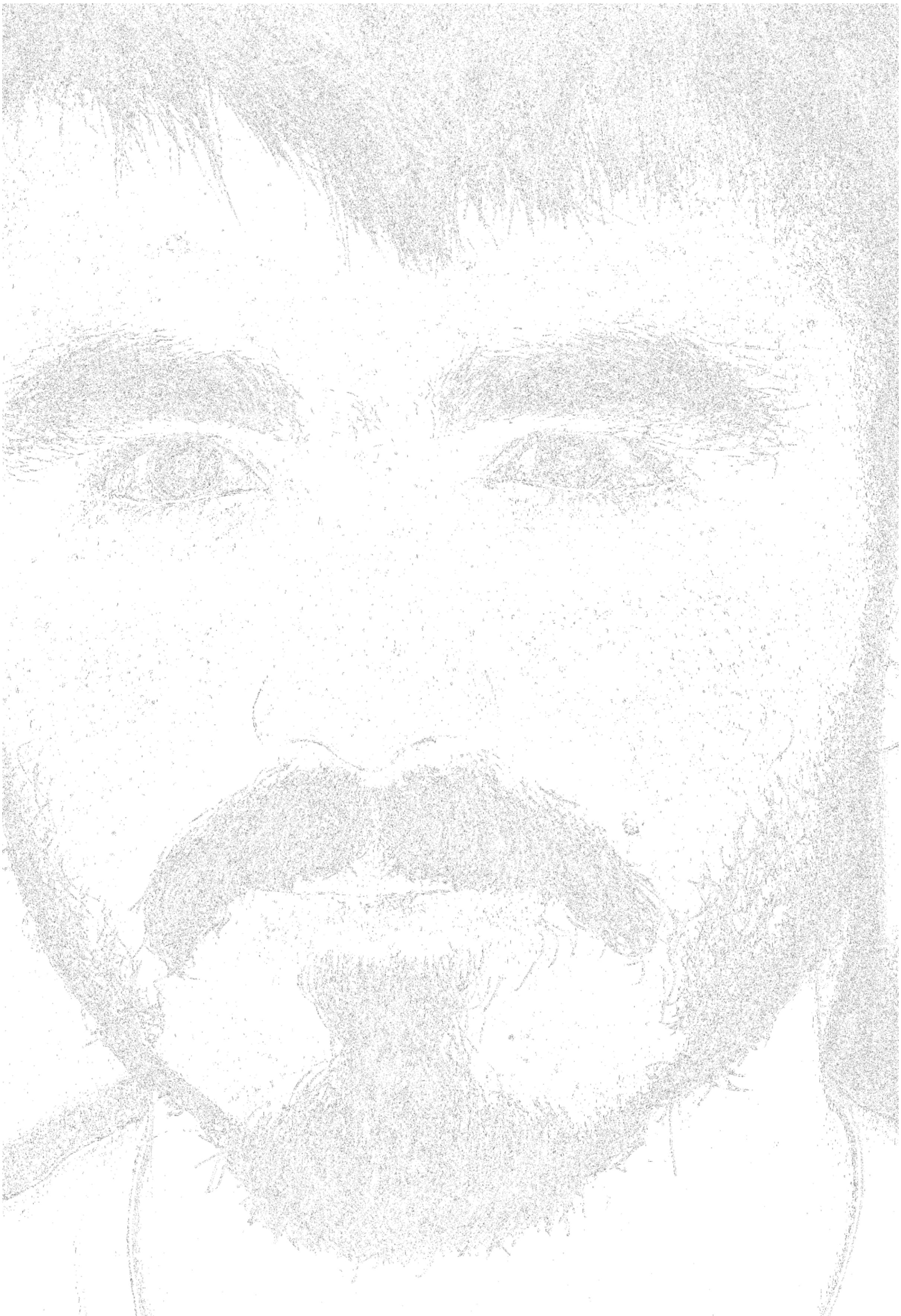


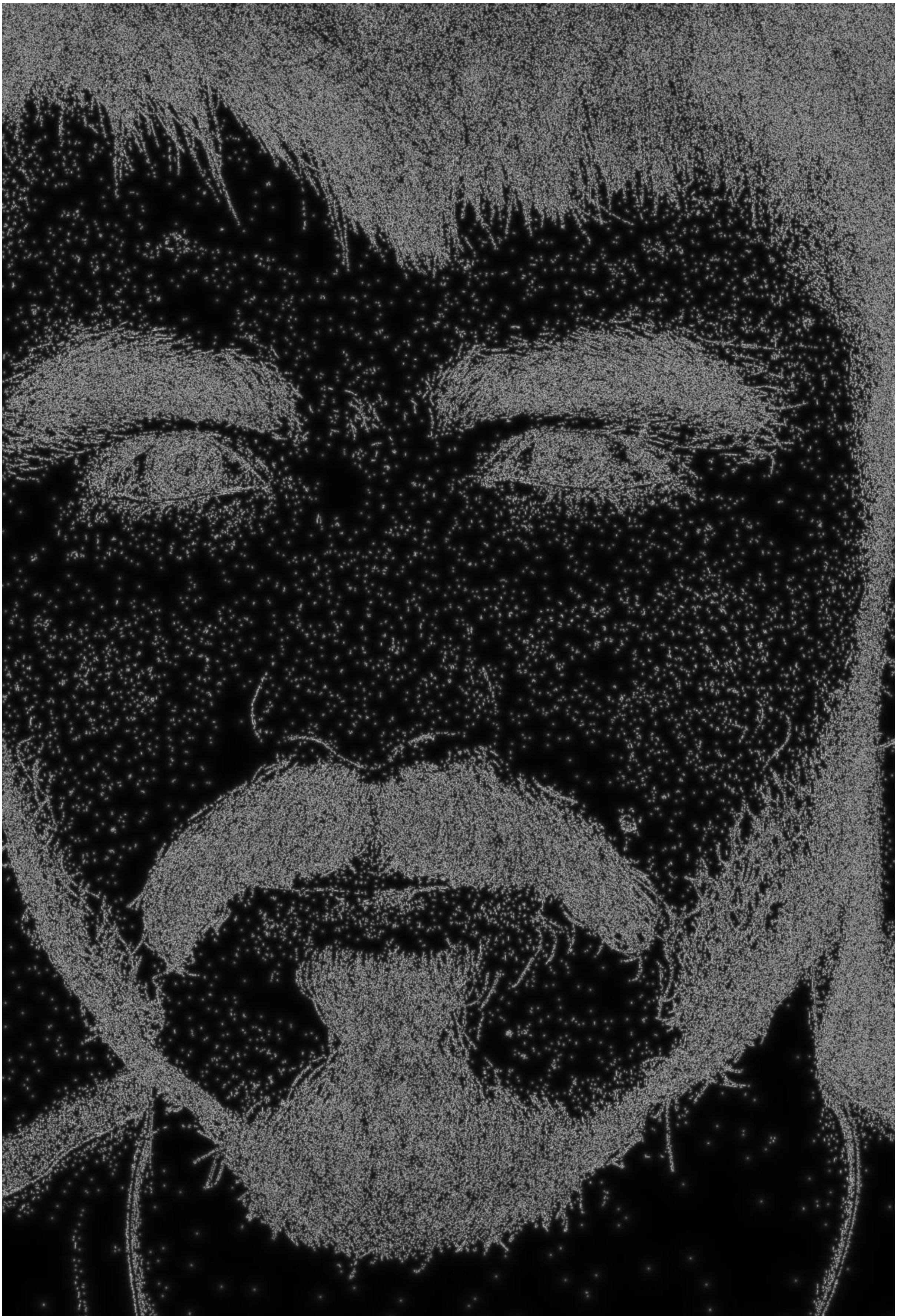
(Original image - 1)



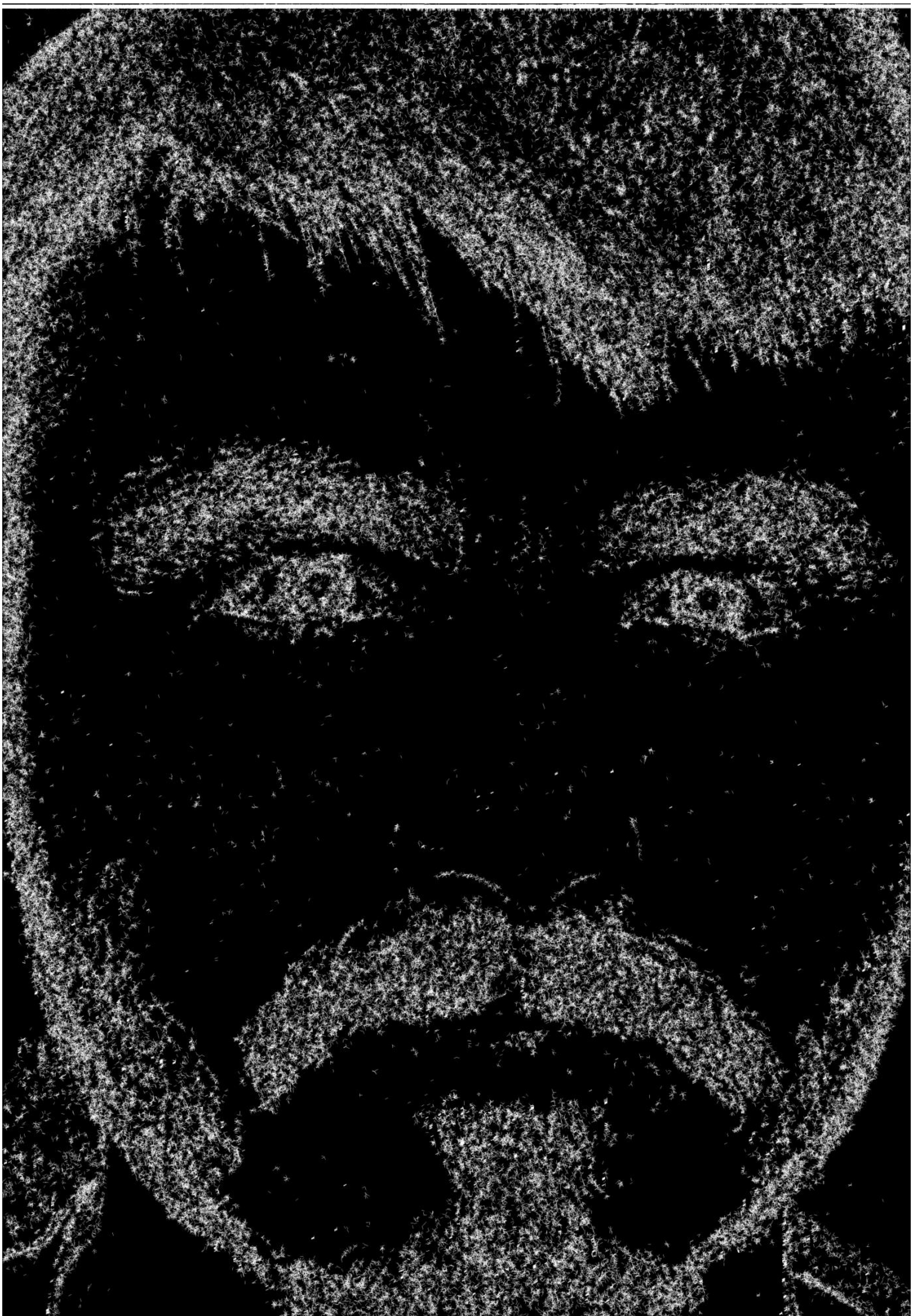




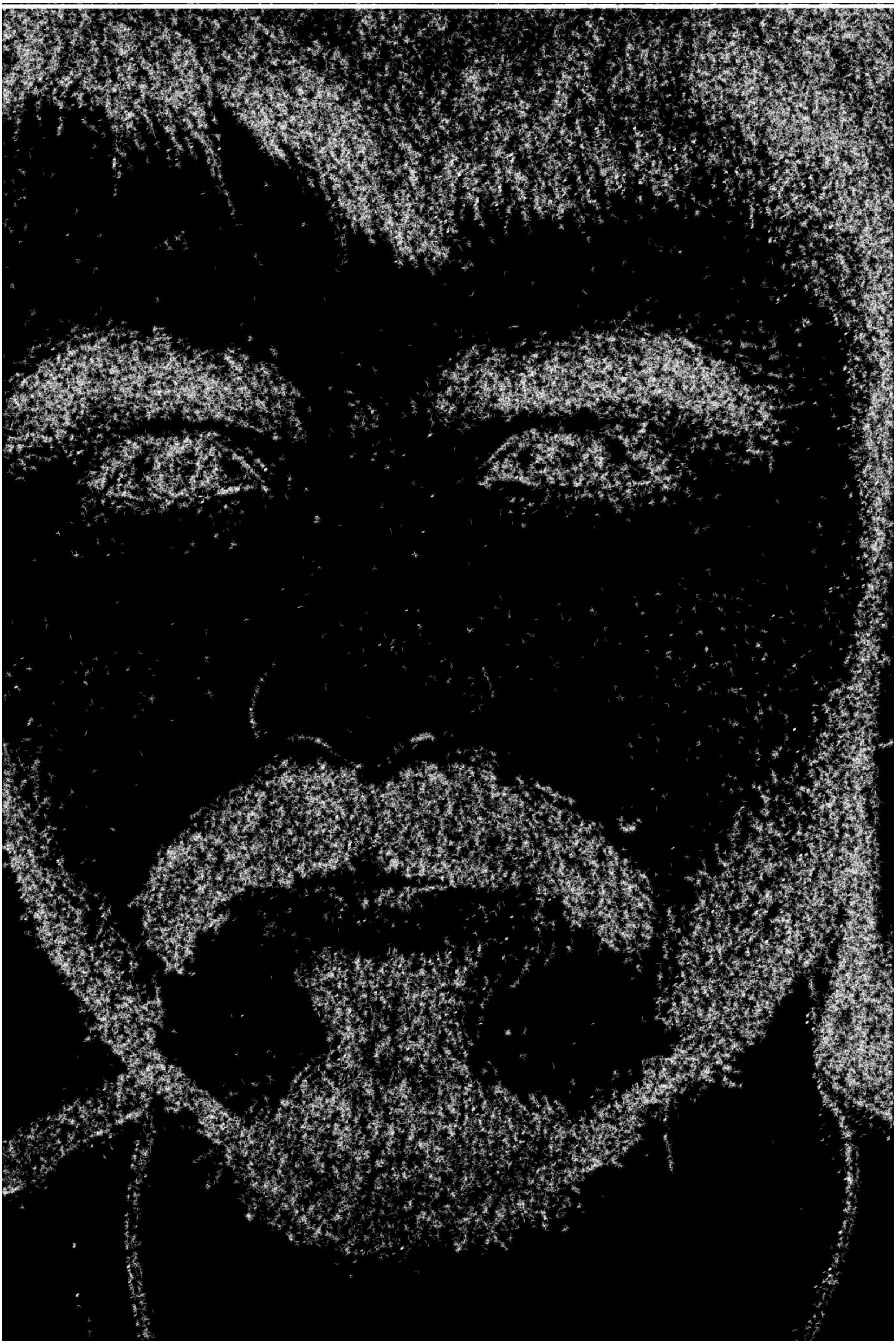




2) Results of growing Hair segments in 2-D:

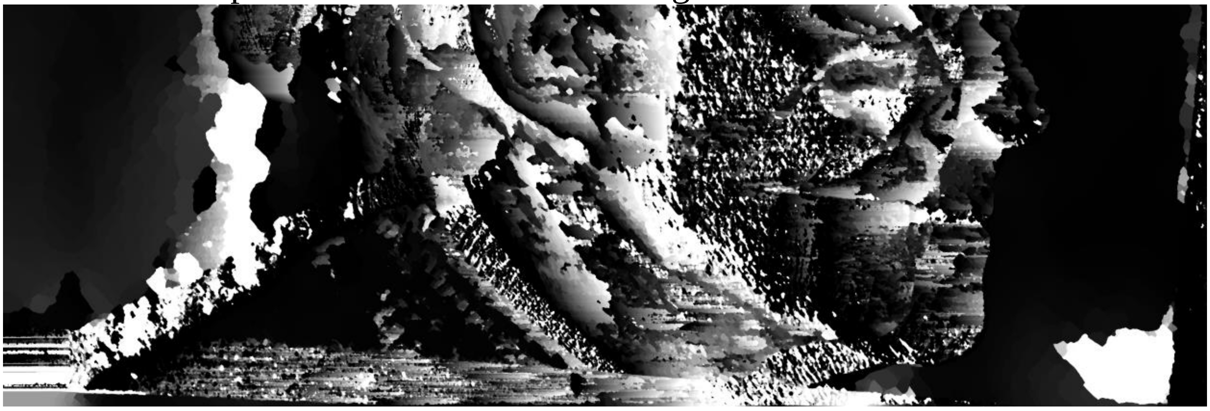


(Grown hair for image-0)



(Grown Hair for Image-1)

2) Generating disparity map for two images. For the generation, a rectification step is needed for which the images are shown.



(Disparity image generated using the next two rectification images)

(The two rectification images used for generation of disparity map image)

Conclusion:

The above is what could be done. The further steps described in the Algorithm section were not implemented. A 3-D reconstruction could not be achieved but hair could be successfully separated and grown in 2-D.

References:

1) Coupled 3D Reconstruction of Sparse Facial Hair and Skin- Thabo Beeler, Bernd Bickel, Gioacchino Noris,
   Paul Beardsley, Steve Marschner, Robert W. Sumner, Markus Gross
   Disney Research Zurich 2ETH Zurich 3Cornell University

2) High-Quality Single-Shot Capture of Facial Geometry - Thabo Beeler Bernd, Bickel, Paul Beardsley, Bob Sumner, Markus Gross