Capture of hair geometry using white structured light

Yin Chen a,c, Zhan Song b, Shuai Lin c, Ralph R. Martin d, Zhi-Quan Cheng b,c,*

a School of Defense Engineering, Army Engineering University of PLA, China
b Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
c School of Computer Science, National University of Defense Technology, China
d School of Computer Science & Informatics, Cardiff University, Wales, UK
* Avatar Science Company, Changsha, Hunan, China

A R T I C L E   I N F O

Article history:
Received 21 December 2016
Accepted 26 October 2017

Keywords:
Structured light
Hair
Avatar
Strip-edge-based coding
Point cloud consolidation

A B S T R A C T

Structured lighting (SL) scanning technology has been successfully applied to reconstruct personalized avatars, providing high-quality geometry for the face and body. However, previous white SL methods have typically been unable to capture and reconstruct hair geometry, preventing a complete, realistic avatar surface from being reconstructed. In this paper, we propose a novel hair capture system, producing a full-head manifold with complete hair, based on four white SL scanners made by ourselves. The key technical contribution is a robust strip-edge-based coding algorithm, using a projection pattern of 18 stripes; it allows geometric acquisition to an accuracy of 1 mm. Experiments are given to show that our system produces high-quality hair geometry, comparable to that captured by more expensive state-of-the-art multiview stereo systems. Our system is however easier to configure, and is more suited to real-world setups.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Work on structured lighting (SL) scanning dates back over 40 years [1,2], and it is one of the most reliable techniques for capturing object surfaces. It has many intrinsic advantages, such as ease of configuration, and suitability for real-world applications outside a laboratory environment. Various SL methods [3–5] have been proposed, using binary or continuous light patterns sequentially projected onto the object of interest, while a digital camera captures images of it. Structured lighting has been widely applied to reconstruct personalized avatars with high-quality geometry for both the face and body. However, to the best of our knowledge, current white SL systems are unable to capture complete avatars, primarily due to the difficulty of capturing and reconstructing hair on the head.

The difficulty in hair reconstruction is mainly caused by the relatively low contrast in reflections, particularly for dark hair. A further problem is that light may penetrate through the hair if it is thin and sparse. 3D range data generated from such reflections is almost always unreliable, resulting in errors and artifacts in the captured geometry. This problem is made worse by the fine geometric detail of the hair: a state-of-the-art commercial SL system can achieve a resolution of about 1 mm, while the average diameter of a single hair fiber is about 0.08 mm. Therefore, it is widely believed that white SL is unsuitable for hair capture.

Fortunately, however, in many applications, the goal is to capture overall hair geometry, rather than precise detail of every strand. This is possible and meaningful as the hairs on the head do not form a random structure, but generally lie in groups of hairs forming tufts and curls. The general method of choice for high-quality hair geometry capture is to use expensive multiview stereo [6] with a complex, carefully configured camera array. However, as Echevarria notes [7], hair reconstruction from multiview stereo is very difficult to achieve outside a laboratory setting. Thus, as in [7], the goal of our technique is to reconstruct a high-quality overall hair surface, unlike other work which aims to reconstruct individual strands of hair [8–13].

In this work, we present a novel white SL hair capture system that produces high-quality overall hair geometry, with results which are comparable to those from state-of-the-art multiview stereo systems having that same goal. Our system can handle various hair-styles ranging from short to long, and from straight to curly. The captured hair geometry can be combined with the rest of the scanned head to build a full-head model (see Fig. 1).

The key technical contribution of the system is a robust strip-edge-based coding algorithm, using a projection pattern of 18 stripe images to provide a geometric acquisition accuracy of better than 1 mm. This algorithm is very reliable, and insensitive to hair appearance in terms of color, texture (e.g. curls) and reflectivity.
Acquisition of hair geometry data using a single white SL scanner we have built around this algorithm takes just 0.3 s. Our overall system uses four such scanners in a laboratory; they capture data sequentially to avoid inference between projected light patterns. Thus, scanning overall takes about 1.2 s, which is a short enough time for a subject to hold steady, with little movement between different scans.

After the raw hair data is captured by our system, erroneous points in regions with sparse and short hair (e.g., the temples) are re-positioned using a point cloud consolidation process, which is our second technical contribution. In such regions, the hair position may appear offset because of light penetration through the sparse or short hair to the scalp. Point cloud consolidation is achieved by a variant of a bilateral filtering scheme which moves the offset points to their expected positions. This allows our system to achieve visually pleasing reconstruction results even in the presence of sparse and short hair.

In the rest of this paper, we briefly review related work in Section 2. We then give technical details of data acquisition in Section 3 and point cloud consolidation in Section 4. Section 5 presents experiments and analyzes the results. Conclusions are discussed in Section 6.

2. Related work

Hair capture is an active research topic in computer graphics. A recent survey [14] has summarized and categorized existing methods from different perspectives. As previously noted, we focus on the smooth surface reconstruction of the overall hair style [7], rather than reconstructing highly-detailed individual wisps or hair strands as is done in [8,9,11,12].

State-of-the-art approaches for overall hair capture are often based on multi-view stereo reconstruction systems, using multiple digital cameras [6]. A typical example can be found in [7], which reconstructs a personal hair-style, with fine-scale geometric detail suitable for e.g. 3D printing. However, as noted by [7], such multi-view systems are costly, and have complex capture setups which are more suited to the laboratory than the real world.

Commodity RGB-D cameras (e.g., Kinect [15], Primesense [16], Xtion Pro Live [17]) provide low-cost scanning hardware, which can be used as a basis for capturing 3D personalized avatars with hair. The pioneering KinectFusion [18,19] is a GPU-based capture system using an RGB-D camera for both tracking and rigid surface reconstruction, allowing users to incrementally capture geometrically accurate 3D models. It has motivated many follow-up algorithms, using differing strategies to improve tracking speed and to expand spatial mapping capabilities to larger scenes or deformable avatars. In particular, the recent DynamicFusion algorithm [20] demonstrates robust performance when reconstructing a non-rigidly deforming human body in real-time. However, such low-cost hardware can only provide low-quality results, with an acquisition accuracy limited to several mm, and further suffers from long capture time due to redundant data collection techniques.

Structured lighting (SL) scanning is the most successfully commercialized technique, for a variety of commercial and technical reasons, including scanning quality, price, reliability, and capture speed. Since the invention of SL scanning [1,2], numerous SL systems have been devised for surface acquisition. Several survey papers [3–5] demonstrate that they have been widely applied to build personalized avatars with high-quality geometry for human faces and bodies. However, the hair on the head of the reconstructed avatar is always missing in previous work, due to the difficulty in distinguishing structured light patterns on dark, highly detailed hair surfaces.

Using expensive near-infrared lights is one approach to providing sufficient contrast in dark areas, allowing the recovery of dark surfaces under general illumination [21]. However, it requires the subjects being scanned to be small and close to the infrared scanner (about 20 mm from it): the limited power of near-infrared light projectors greatly limits the size of the working volume. This limitation restricts its suitability for human hair capture.

Instead, we revisit the question of whether (dark) hair can be effectively captured by using white structured light. The key question is whether the projected signal, in the form of a light pattern, can be coded in a way that can be accurately recovered. As Geng [5] notes, existing SL techniques can be classified into multiple-shot (sequential) and single-shot methods, according to the number of projected patterns. Multiple-shot methods often generate more reliable and accurate results, provided that the 3D target is static. We use a robust, multiple-shot, strip-edge-based coding algorithm to acquire hair geometry. Compared to using raw
image intensities and thin image lines, which are vulnerable to problems arising from the reflective nature of hair surfaces, edges of binary strips in the illuminated pattern are more localizable and better preserved in the image data. Higher localizability of the edges leads to more accurate reconstruction and higher robustness of our SL scanning system.

The original strip-edge-based SL framework [22] was intended to handle natural reflective surfaces, e.g. shiny coins and metallic workpieces. The improvements in our algorithm over that work are twofold. Firstly, we use a more robust and reliable strip-edge-based coding algorithm, with a different projected pattern. Secondly, we achieve a faster capture speed by decreasing the number of projected pattern images from the original 30 to 18, leading to a scanning time of about 0.3 s. Our experiments demonstrate that the resulting algorithm can effectively reconstruct a full-head manifold for various hairstyles, which cannot be achieved by the original strip-edge-based SL framework [22].

3. Point acquisition

We first overview our approach to data acquisition in Section 3.1, then present our optimized strip-edge-based coding algorithm in Section 3.2. Configuration of our hair capture system is described in Section 3.3.

3.1. Structured light scanner approach

Our white SL scanner has three main components, as shown in Fig. 2(a): a white light Benq MX620ST projector with a resolution of $1024 \times 768$ pixels and frame rate of 60 Hz, a CMOS PointGray camera using a Fujinon zoom lens, with resolution $2048 \times 2048$ pixels and frame rate of 90 Hz, and an FPGA controller which synchronizes the projector and camera.

The distance between the camera lens center and the projector is about 0.3 m, enabling use of an accurate wide baseline calibration method [23]. The focal lengths of projector and camera are both set to 1.6 m.

The scanning volume of the SL scanner is a rectangular frustum, where the near and far rectangles of the truncated pyramid are $0.81 \times 0.61$ m and $1.19 \times 0.89$ m, respectively, at distances of 1.3 m and 1.9 m. This means that the central rectangle is $1.0 \times 0.75$ m, at a distance of 1.6 m. As there are 1024 projected pixels along each horizontal line, the acquisition resolution of the system is 1.0/1024 m or just less than 1 mm.

As Fig. 2(b) shows, the controller sends binary light patterns to the projector. These are projected onto the subject surface and imaged by the camera. Correspondences between projected and recovered patterns are found and used by the attached computer to extract 3D point information.

3.2. The strip-edge-based coding algorithm

The strip-edge-based coding algorithm projects a sequence of patterns of black and white vertical strips; the camera acquires an image for each projected pattern, and the strip edges in the image data are precisely located. Each strip is determined to be black or white in any one image, giving one bit of information; the bits from successive images build up a unique binary code for each strip, which differs from the code of any other strip. As in [22], the unique codeword for each binary point is a combination of a global pseudo-Gray code and a local strip-edge code. Correspondences between the illumination pattern and the image plane are established, allowing spatial depth at the strip-edge points to be determined from the binary code.

A significant difference between our algorithm and the original algorithm in [22] is the design of projected patterns, especially we propose a new sequence of projected patterns and decrease the number of patterns from 30 to 18. This is based on careful analysis of the strip-edge-based coding scheme, and an observation that it was somewhat redundant in [22]. A direct benefit of this reduction is to reduce scanning time from 0.5 s to 0.3 s (as the projector frequency is 60 Hz, the total scanning time is 18/60 = 0.3 s). This is fast enough for a human subject to hold a more or less static pose during hair acquisition.

Fig. 3 illustrates our sequential projection patterns for the 18 images with a resolution of $1024 \times 768$ pixels. In a strip-edge-based coding scheme [22], the first 1–10 images form a pseudo-Gray code sequence, then the final 8 images use a shifting-strip pattern to precisely locate edges. First two images are mainly used to determine reference intensity levels, while images 3–10 gradually divide the whole region with a sequence of successively finer patterns. The pseudo-Gray coding scheme in this way constructs 128 vertical subregions of width 8 pixels, each with a unique codeword. Rather than simply using more images in the pseudo-Gray coding scheme to get 1024 subregions, each corresponding to a column of pixels, instead, we use 8 further images based on shifting strips to accurately locate edges—this allows precision to sub-pixel level, as will be explained. The last 8 images consist of alternating black and white strips, each of width 4 pixels; each image is relatively shifted right by 1 pixel.

The original strip-edge-based SL framework [22] was intended to handle natural reflective surfaces such as shiny coins and other metallic workpieces. However, our goal is to capture black hair, which necessitates three major differences in our approach for two such different materials.

(1) The first filtering scheme for the all-white and all-black images has experimental thresholds. Ours are experimentally determined using the black image, while [22] uses a default value throughout the program’s execution. It is clear that the thresholds affect the final results, since the following steps use them to mask out pixels which do not give reliable depths.

(2) Our pattern is different from the one in [22]. The latter uses a code subdividing the 1024-pixel image width into 256 subregions, each of width 4. Then a strip pattern of width 2 is shifted 3 times in steps of 1 pixel (giving 4 images including the initial pattern), which is used to encode each subregion with additional bits. Instead, our paper uses Gray code to subdivide the 1024-pixel image width into 128 subregions, each of width 8. A strip pattern of width 4 is shifted 7 times in steps of 1 pixel (giving 8 images including the initial pattern). We increase the width of the strip pattern from 2 to 4, the method in [22] is aimed at shiny surface micromeasurement, and in this case, smaller subregions lead to less ambiguity. Also, as the object is close to the camera, the black and white border of the strip pattern is easy to recognize. Our paper targets hair capture, and does not need high resolution. We thus increase the width of the strip pattern for two reasons. Firstly, as the hair is far from the camera, using a wider strip pattern makes it easier to recognize the black and white border of the strip pattern. Secondly, the light strip is more easily absorbed by black hair, and if it is too narrow, it cannot adequately reflect the differences in brightness under light and dark strips.

(3) This approach decreases the number of patterns from 30 to 18, which reduce scanning time from 0.5 s to 0.3 s. This makes it easier for a human subject to hold a static pose during hair acquisition.

We now describe how the method works. We use the notation $P_j$ for projected image $j$, and $C_j$ for captured image $j$.

Firstly, an all-white image $P_1$ and an all-black image $P_2$ are projected in turn, $C_1$ and $C_2$ are captured, and a contrast image $I_{1,2} = (C_1 - C_2)$ is obtained by calculating the intensity difference of the two captured images. Certain pixels are likely to be unreliable, and should be ignored. A lower threshold $T_1$ and an
Fig. 2. White structured light scanner setup. (a) The hardware includes a projector, a calibrated camera with a zoom lens, and an FPGA controller. (b) The controller sends binary light patterns to the projector. These are projected onto the subject surface and imaged by the camera, allowing 3D point information to be extracted by the computer.

Fig. 3. The 18 strip-edge-based patterns with resolution 1024 × 768 pixels which are sequentially projected onto the subject.
upper threshold $T_2$ are set separately for $l_{1,2}$. Empirically, we set $T_1 = 5$ and $T_2 = 245$, based on experiments. If the difference value at a pixel is less than $T_1$, it is in shadow, or is a strongly reflecting (highlight) region, so unreliable. Such pixels are masked out in the following steps, as they do not give reliable depths. Similarly, if the $T_1$ region, so unreliable. Such pixels are masked out in the image $P_{11}$.

The image $P_{11}$ is the conjugate (inverse) of $P_{13}$ (see Figs. 3 and 4), with black and white strips swapped. In the same way, image $P_{13}$ is the conjugate of $P_{11}$, etc. These image pairs are then coded as follows, taking images $C_{11,15}$ and $C_{15,15}$ as a specific example.

Consider the projected images in a conjugate pair. The contrast image $P_{11,15}$, $P_{11,15} = P_{11} - P_{13}$, results in an intensity image having three values: $-255$, $0$, and $255$, shown as Fig. 4(a), where the value $0$ only occurs at the boundary edge between two strips. Such boundaries can be further divided into two types: in the first, $-255$ switches to $255$, while in the second, $255$ switches to $-255$. In our coding scheme, the first case is coded as $0$, and the second case as $4$. The coding scheme handles the other three conjugate image pairs in a similar way: $P_{12}$ and $P_{16}$ provide codes $1$ and $5$, $P_{13}$ and $P_{17}$ provide codes $2$ and $6$, $P_{14}$ and $P_{19}$ provide codes $3$ and $7$. Consequently, these $8$ cases can be coded as values in $[0, 7]$ in each of the $128$ subregions.

When the projected image is captured, the locations of boundary edges between white and black strips are less well defined, for a variety of reasons including the optical properties of the camera, the nature of the reflection from the subject’s surface, material properties, and so on. These cause the projected strip edges to be somewhat blurred in the captured images as schematically shown in Fig. 4(b). To overcome this problem, we locate the boundary edge by data fitting. Empirically the data shows intensity variations similar to a sine wave, and this can be used to give the location of the boundary edge of each strip [22]. The strip edges can be determined to sub-pixel accuracy by zero crossings in the contrast image $C_{11,15} = C_{11} - C_{15}$, assuming a sine wave is a good model. We divide $C_{11,15}$ into two segments, $C_{11,15}^+$ and $C_{11,15}^-$, for which $C_{11,15}^+ > 0$ and $C_{11,15}^- < 0$, respectively. The zero-crossing positions of $C_{11,15}^+$ and $C_{11,15}^-$ are determined by first finding where the second derivatives $\nabla^2 C_{11,15}^+$ and $\nabla^2 C_{11,15}^-$ are zero. The strip-edge position $x_{11,15}$ is then estimated to be at the average of these two locations.

Finally, the 3D point information is computed based on well-known SL triangulation principles [3,4].

3.3. Configuration of hair capture system

Using four scanners of the kind described, we have built an experimental system to capture hair, as shown as Fig. 1. The four scanners point inwards from the vertices of a square, whose semidiagonal distance is $1.6$ m. Face capture is important, so one scanner is located at a lower height, directed towards the subject’s face. The other three, located a little higher, are inclined downwards to better capture the hair on the subject’s head.

The scanners capture data sequentially to avoid interference between the projected patterns. As each scanner takes $0.3$ s to capture data, the whole scanning time is $1.2$ s.
Fig. 5. Depth discontinuities occur in certain regions in left and right camera views, due to measurement of scalp as well as hair.

Fig. 6. Consolidation of a left-side partial point cloud: (a) original data, (b) result of standard bilateral filtering, (c) our result, before and after post deformation computation. Note position changes for the erroneous hair regions marked in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 7. Depth correction pipeline: (a) projection, (b) erroneous hair region detection, (c) generation of indicator images, (d) sorting of erroneous hair points and filtering, (e) post deformation.

The 3D point cloud data produced from the four scanners is registered using the method in [24] to put all 3D coordinates into a common reference frame.

4. Point cloud consolidation

We now use a point cloud consolidation technique to help fix erroneous points in regions with sparse and short hair (e.g. the temples). Its main goal is to change the position of points which lie on the scalp to a position more suitable to represent the hair. This allows us to successfully reconstruct a suitable hair surface.

4.1. Approach

Our basic hair capture system can successfully tackle most hair scanning problems. However when capturing sparse (e.g. crew-cut style) hair, positional errors can occur, particularly in regions with short hair (e.g. the temples): a typical example is shown in Fig. 5. The main problem is the sparsity of the hair, which does not form a sufficiently dense and continuous region to reflect the light emitted by the projector: light penetrates through the short hair onto the scalp, whose position is measured instead. Constructing a surface which within a local region is sometimes determined by hair, and sometimes scalp, is undesirable. We would like to
We can thus regard points lying on the scalp as being erroneous, as they have a positional offset from their desired locations on the notional hair surface. We adjust them by means of a modified bilateral filtering approach.

In standard bilateral filtering [25], the depth value at each pixel is iteratively replaced by the weighted average of depth values of neighboring pixels. Each pixel’s depth value is updated immediately but propagated to nearby pixels in the subsequent iteration. However, our experiments demonstrate that this standard strategy is unsuitable for the current problem, leading to incorrect smoothing of the erroneous regions, as shown in Fig. 6(b). The main reason is that each pixel merely affects its local neighbors in the standard approach.

We thus modify the bilateral filtering approach using the following principles during each position updating iteration. Firstly, each point being updated should only be influenced by those neighbors which have already been updated in this iteration or
have been marked as correct and having fixed position. Secondly, once updated, the changed positions should be propagated immediately to any subsequently updated points in the current iteration. The consequence is that if filtering considers erroneous hair points in the correct order, information from correct points is rapidly propagated to incorrect points. This leads to better results, as shown in Fig. 6(c).

4.2. Implementation

Success of this revised bilateral filtering approach relies heavily on correct detection of erroneous regions, and use of an appropriate ordering when updating these points. The idea is implemented as illustrated in Fig. 7.

Firstly, each head point cloud is projected back to four depth images, in the viewports of the left, right, front and back scanners. Projection is performed by using the calibrated positions of the three scanners. We now consider the left side case as an example to explain the approach.

Erroneous hair regions are determined with manual assistance. Segmentation is performed on the 2D intensity image captured by the left scanner. The segmentation is then transferred to the depth image, using the pixel-to-pixel correspondence between them.

Initially, the user (approximately) manually segments the image into two areas: the hair area, and all else, including face, neck, etc. Manual segmentation is performed with the 2D image GrabCut tool, as in [8]; it requires just a few strokes to be drawn. Next, within the hair region, correct hair points are manually selected by painting the 3D point cloud using a stroke. Any remaining hair area is considered to potentially be erroneous. (If no correct hair points are selected, data consolidation is skipped.)

Information from the above step is then stored in two indicator images attached to the depth image. The first is a binary validity image, which indicates whether a pixel is valid or not; where small holes exist in the 3D-to-2D projection depth image, the missing data is invalid. The second image is a correctness-weight image. Its pixels are set to one of three values, and indicate whether a pixel is erroneous or not. If the underlying hair point is erroneous, its value is set to 0. If the underlying point lies outside the hair area (e.g. face, neck), its value is set as 1. If the underlying hair point is correct, its value is set to α, empirically set as 5. Since the manually selected hair region is more reliable, we give a larger correctness weight (of 5) to pixels in this region. Pixels neighboring the underlying hair region will have larger values after convolution, so will be processed first.

The position of each erroneous point is now iteratively updated using the revised bilateral filtering process. In each iteration, each point is processed in two steps:

- Sorting. All erroneous points to be processed are ordered using a weight convolution scheme. The validity and weighted correctness indicator images are both convolved with a local discrete 9 × 9 Gaussian distribution, and two scalar values Vvalid and Vcorrect are computed for each erroneous point. The convolved weight value is then set to Vcorrect/Vvalid + ϵ (ϵ experimentally set as 0.1), for each depth pixel corresponding to an erroneous point. The erroneous points are sorted in descending order of convolved weight.
- Revised bilateral filtering is performed with position updating and weight propagation. The depth value of the point is updated, using standard bilateral filtering [25]. The processed point is removed from the processing queue.

The above scheme ensures that the appropriate constraints are taken into account during the revised bilateral filtering process, which utilizes a dynamically sorted queue. The width of the smoothing window used in bilateral filtering is Wspatial = 4 pixels, while the range weight is Wr = 0.02. The erroneous points are processed once time within one pass. Iteration terminates when no erroneous points remain.

Finally, the hair details are recovered by use of a post deformation scheme. The bilateral filtering process results in hair regions becoming over-smoothed, with lost details. To recover the original captured detail, the embedded graph method [26] is used to deform the original surface to the position-smoothed surface. Graph nodes are distributed according to each vertex’s depth gradient; higher gradient vertices have a reduced probability of being assigned as graph nodes, with the intent of keeping more local rigidity. The effect of the post deformation scheme is illustrated in Fig. 6(c).

Note that, for a real subject, a 3D raw point cloud was initially produced by the acquisition technique in Section 3, then data consolidation (see Section 4) was optionally used when deemed necessary by the user. The final surface was obtained by the screened Poisson reconstruction method [27].

5. Results

We now evaluate our capture system for effectiveness and robustness on different hair styles, and make a comparison with other systems and methods. We also consider its limitations.
5.1. Evaluation

The system has been validated by capture and reconstruction of personalized avatars for many human subjects. Typical results are shown in Figs. 1 and 8, for subjects with a variety of hairstyles ranging from short to long, from straight to curly. As the top example in Fig. 8 shows, even short black crew-cut hair style can be plausibly captured. While this case was optimized by the data consolidation process, it was not necessary for the example in the second row as the hair in the temple regions was denser, nor was it needed for the bottom two examples. In the third row, there is some banding visible in the long straight hair, but we note this is actually present in the subject’s hair, which had been given a permanent wave, and is not an artifact caused by our approach. Fig. 9, shows a challenging ‘Afro’ hair style. Our system can also reconstruct a satisfactory result in this case.

The capture process takes 1.2 s. Using a laptop PC with a 1.6 GHz Intel Core i7 processor, about 60 s are spent for data processing, of which 20 s are needed for the generation of the raw point cloud, and 40 s for reconstruction of a surface with between 300 and 600k vertices. The time taken for point cloud consolidation is determined by the size of any erroneous hair regions; revised bilateral filtering takes about 3 s for one region.

5.2. Comparison

The reconstructed surfaces contain rich geometric details specific to the subject, with a resolution of about 1 mm. A quantitative evaluation has been made by using a template, captured both by our white SL system and using a state-of-the-art multiview stereo system. Fig. 10 validates our reconstructed hair against an accurate template of the mannequin produced by the multiview stereo system [6]. It comprises 20 Canon 700D SLR cameras placed in a quarter-sphere, each image having a resolution of $5184 \times 3456$. We reconstruct point clouds from 5 different views (front, back, left, right, top) and align them using ICP [28]. The system has an accuracy of about 0.1 mm. For comparison, gel was used on the mannequin hair, to gather hair into a fixed hairstyle with well
defined clumps for capture by both the multiview stereo system and our white SL system. Hair with gel was better captured than plain hair in both systems, mainly because of the effectiveness of gathering the hair into clumps. The $L_2$ distance error between the reconstructed mesh and the template is shown on the right. The maximal distance error is about 1.0 mm, and the average error is about 0.7 mm, in agreement with our expectation that our method should provide results with an accuracy of just under 1 mm.

Thanks to the novel design of the projection pattern in the SL framework, our strip-edge-based coding algorithm produces better results than the original approach in [22]. As shown in Fig. 11, the latter is unable to produce a hair model. An additional advantage of our algorithm is that our scanning speed is also faster.

5.3. Limitations

While our system can handle a wide range of hairstyles, it has problems with thin and sparse hair: see Figs. 5 and 12. Artifacts arise as light penetrates through the thin and sparse hair. As the system resolution is about 1 mm, but the average diameter of a single hair fiber is 0.08 mm, individual hairs, or small groups of hairs are not treated separately. Our system produces more visually pleasing results for thicker hair styles.

Due to the need to capture multiple images, our binary coding algorithm is unsuited to moving subjects. Higher frequency projectors and cameras could shorten the capture time; using a smaller number of binary patterns could also accelerate the capture speed.

During point cloud consolidation, manual selection of erroneous hair regions must still be performed. Although the stroke-based tool is easy to use, it would be preferable to provide an automatic detection method. This could be based on detecting rapid changes in depth.

Our method just reconstructs hair geometry without color. A seamless texture composition method [29] can be used to add color to the reconstructed model.

There is more noise in the marginal areas of the captured subject. While they can always be captured by two systems in the setup, we can use the relative position of the two systems to remove some noise in the common area. Furthermore, we can make use of the color information to enhance the reconstruction quality.

Fig. 11. Comparison with an earlier SL approach. Left: photograph of the subject. Center: our result. Right: the method in [22] fails to capture hair.

Fig. 12. Limitations. For thin and sparse hairstyles our system may fail to generate a reliable result; geometric details are lost in the red areas. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

6. Conclusions

We have proposed a novel hair capture system which can produce plausible hair geometry with rich details. This appears to be the first approach to solving the challenging hair capture problem using white structured light. The basis of the approach is a robust strip-edge-based coding algorithm, with an acquisition resolution of 1 mm. The system is low-cost, easy to implement, and shows good potential for personalized 3D avatar modeling.

Acknowledgments

This work was supported by the Natural Science Foundation of China (No. 61602507), Natural Science Foundation of Jiangsu Province (No. BK20150723), Key Researching plan (International and regional cooperation) of Hunan Provincial Department of science and technology (No. 2016WK2038), China Postdoctoral Science Foundation (No. 2016M602555), National Key Research and Development Program of China (No. 2017YFB1103600), Shenzhen Science Plan (JSGG20150925164740726), and the National Natural Science Foundation of China (Nos. 61375041, U1613213).
References


