Computers & Graphics 59 (2016) 28–38

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

Automatic garment retexturing based on infrared information ☆

Egils Avots [a], Morteza Daneshmand [a], Andres Traumann [b], Sergio Escalera [c], Gholamreza Anbarjafari [a,d,*]

[a] iCV Research Group, Institute of Technology, University of Tartu, Tartu 50411, Estonia
[b] Institute of Computer Science, University of Tartu, Tartu 50409, Estonia
[c] Computer Vision Center, Universitat de Barcelona, Spain
[d] Department of Electrical and Electronic Engineering, Hasan Kalyoncu University, Gaziantep, Turkey

ARTICLE INFO

Article history:
Received 21 November 2015
Received in revised form
4 May 2016
Accepted 6 May 2016
Available online 19 May 2016

Keywords:
Garment Retexturing
Texture Mapping
Infrared Images
RGB-D Acquisition Devices
Shading

ABSTRACT

This paper introduces a new automatic technique for garment retexturing using a single static image along with the depth and infrared information obtained using the Microsoft Kinect II as the RGB-D acquisition device. First, the garment is segmented out from the image using either the Breadth-First Search algorithm or the semi-automatic procedure provided by the GrabCut method. Then texture domain coordinates are computed for each pixel belonging to the garment using normalised 3D information. Afterwards, shading is applied to the new colours from the texture image. As the main contribution of the proposed method, the latter information is obtained based on extracting a linear map transforming the colour present on the infrared image to that of the RGB colour channels. One of the most important impacts of this strategy is that the resulting retexturing algorithm is colour-, pattern- and lighting-invariant. The experimental results show that it can be used to produce realistic representations, which is substantiated through implementing it under various experimentation scenarios, involving varying lighting intensities and directions. Successful results are accomplished also on video sequences, as well as on images of subjects taking different poses. Based on the Mean Opinion Score analysis conducted on many randomly chosen users, it has been shown to produce more realistic-looking results compared to the existing state-of-the-art methods suggested in the literature. From a wide perspective, the proposed method can be used for retexturing all sorts of segmented surfaces, although the focus of this study is on garment retexturing, and the investigation of the configurations is steered accordingly, since the experiments target an application in the context of virtual fitting rooms.

1. Introduction

Retexturing constitutes an essential step towards creating models providing realistic visual representations of real-world objects, i.e. image synthesis [1–4]. In fact, it is intended to incorporate colour information into a model, which may be the result of a 3-Dimensional (3D) reconstruction process [5]. Usually, performing the latter task involves mapping a planar texture image, which contains the desired colour pattern, onto the surface of the object, being referred to as texture mapping [6]. Various techniques have been proposed in the literature for fulfilling the foregoing goal. An intermediate 3D shape [7], direct drawing onto the oct [8,9] and using an exponential fast matching method through making use of geodesic distances [10–12] could be mentioned as examples.

Among the most important applications of retexturing is its usage in 3D virtual garment representation, the realistic appearance of which is vital, and has encouraged many researchers to get engaged in attempts to improve it during the last decade [13–15]. The role it plays in movie and game industries [16,17] is also of paramount importance. One of the most frequently used texture fitting methods was proposed in [18], allowing the users to sketch garment contours directly onto a 2-Dimensional (2D) view of a mannequin. The initial algorithm has later been further enhanced by others [19,20]. An alternative approach to this problem is using a single image. In [21], an estimation of a 3D pose and shape of the mannequin is followed by constructing an oriented facet for each bone, according to their angles, and projecting the 2D garment outlines onto the corresponding facets.

In [22], on the other hand, the focus is on a texture mapping algorithm on the basis of harmonic maps. They project the 3D surface onto a plane, and parametrise it, by means of an *angle-based-flattening* method. The constraints between the feature-points of the model and the texture image are then specified by the user interactively. Afterwards, the texture coordinates are determined using harmonic maps. Additionally, for achieving reliable real-time performance, the whole map is locally refined. The resulting algorithm has been reported to be

☆This article was recommended for publication by Kai Xu.

* Corresponding author at: iCV Research Group, Institute of Technology, University of Tartu, Tartu 50411, Estonia.
E-mail addresses: ea@icv.tuit.ut.ee (E. Avots), md@icv.tuit.ut.ee (M. Daneshmand), andres.traumann@ut.ee (A. Traumann), sergio@maia.ub.es (S. Escalera), shb@icv.tuit.ut.ee (G. Anbarjafari).

http://dx.doi.org/10.1016/j.cag.2016.05.002

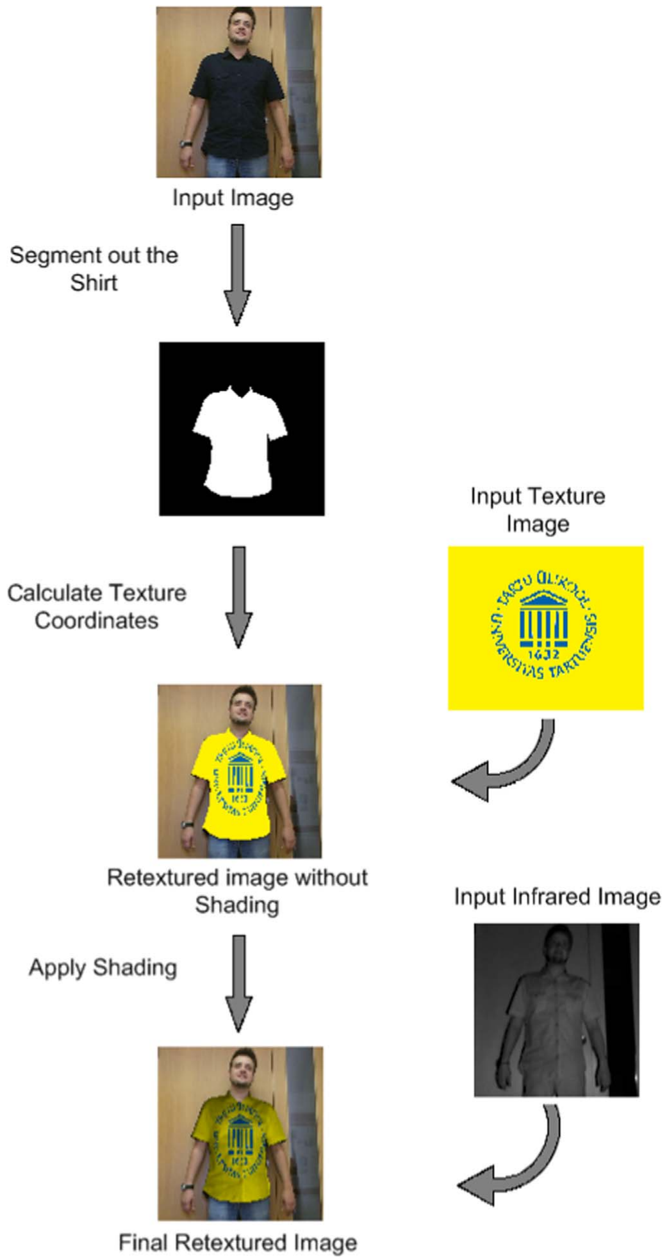

Fig. 1. A Flowchart illustrating the necessary steps of the IRT.

accurate, due to the energy-minimisation embedded into the harmonic maps, which prevents possible distortions.

Retexturing usually involves several challenges. First, texture mapping is difficult in the presence of non-rigid and easily transformable surfaces, such as clothes. One of the main issues to be tackled when dealing with such surfaces is self-occlusion, where a part of the surface blocks the visibility of another one [23]. An additional challenge is to shade the reference texture [24]. Various computer graphics rendering methods could be employed to visualise the surface [25]. However, the lighting intensity and the original colour of the surface are usually not known from the outset, which means that the shading parameters, which are dependent on a map transforming the surface colour to the one reported on the RGB channels, are not available, unless a considerable amount of manual contribution is provided by the user, in non-automatic settings. The foregoing issue is the main reason causing the existing texture mapping methods to degenerate under the changes in the illumination conditions, along with other factors, such as the complexity of the scene.

This paper is aimed at proposing, and verifying the efficiency of, a new approach, referred to as the IR-based Retexturing Method (IRT), which overcomes the above problem, and offers reliable, automatic performance. The underlying notion and the main novelty of the IRT lies in the shading process, which is completely colour- and texture-invariant. In fact, the shading parameters, in the context of the IRT, are extracted from a linear transformation mapping the Infrared (IR) image to the RGB one. More clearly, the pure colour values are derived from the IR image, and, for guaranteeing that the changes of the illumination and other experimental factors are incorporated into the calculations properly, the factors supposed to have been applied to them to produce the RGB representation are extracted, and applied to the reference texture.

The state-of-the-art methods often perform satisfactorily only in cases where the retextured region is sufficiently bright, and its luminance values can be readily utilised. The IRT, however, is able to handle the general case. Furthermore, it operates properly for garments that already have textures on them, which, to the authors' knowledge, is not possible by any of the existing approaches. It is worth noticing that despite the fact that the IRT, based on the experimental results reported in the upcoming sections, is capable of overcoming occlusions and self-occlusions in some cases, these issues are not taken into account in this study, i.e. it is assumed that the garment image does not suffer from such deficiencies.

The IRT has been developed with a focus on virtual fitting room applications [26]. It replaces the texture of a shirt obtained using the Microsoft Kinect II camera with a new custom texture from an image file. As the first step, the image is segmented in order to extract the region standing for the cloth, using two main approaches: Breadth-First Search (BFS) [27] and GrabCut [28]. The latter is intended to minimise the amount of manual contribution required from the user, which will be discussed in more detail throughout the upcoming sections. Then texture mapping is conducted, based on the normalized coordinates of the reference texture image and that of the surface to be retextured. Finally, the aforementioned novel shading approach is incorporated in order to integrate the changes in the illumination and other experimental settings that might have affected the appearance of the surface in the scene. For the sake of evaluating the efficiency of the proposed method, it is examined based on Mean Opinion Score (MOS) analysis, which has resulted in superior scores.

The rest of the paper is organised as follows. Section 2 details the IRT. Section 3 provides a literature review outlining the differences between the IRT and its existing state-of-the-art counterparts. Section 4 presents the results of implementing the IRT under various experimentation scenarios. Finally, Section 5 concludes the paper.

2. Description of the proposed method: IRT

The IRT uses static images from the Microsoft Kinect II camera, along with the depth information and real-world coordinates it provides. Although it is devised such that the most general applications could be handled, the tests are carried out on images that contain a person wearing a shirt. The main application of the IRT is in virtual fitting rooms, where the subject can have the texture of the 3D model of the garment they are wearing changed as they wish.

The IRT replaces the texture of the shirt with a desired reference texture, through performing the following three tasks: segmentation, retexturing and shading. In other words, the part of the image corresponding to the garment worn by the subject is first segmented out, and then retextured by calculating the texture-domain coordinates for each pixel in the area of interest, followed by applying shading to the colour information. The overall step-by-step procedure is shown in Fig. 1, and for more clarity, through the pseudo-code provided in Algorithm 1.

Algorithm 1. A pseudo-code representing the IRT

Input: ColourImage, IrImage, TextureImage, CoordinateMapper, IrLow, IrHigh
Output: RetexturedImage

$SegmIm \leftarrow segmentImage(ColourImage)$
$W \leftarrow ColourImage.width$
$H \leftarrow ColourImage.height$
$IR_{min} \leftarrow \min\{IrImage(x, y) \ : \ SegmIm(x, y) = 1\}$
$IR_{max} \leftarrow \max\{IrImage(x, y) \ : \ SegmIm(x, y) = 1\}$
foreach $(x, y) \in IrImage$ **do**
  $IrImage(x, y) \leftarrow IrLow + \frac{IrHigh - IrLow}{IR_{max} - IR_{min}} \cdot (IrImage(x, y) - IR_{min})$
foreach $(x, y) \in ColourImage$ **do**
  $i \leftarrow 0$
  if $SegmIm(x, y) = 1$ **then**
    $(X, Y, Z) \leftarrow CoordinateMapper(x, y)$
    $XCoords[i] \leftarrow X$
    $YCoords[i] \leftarrow Y$
    $i \leftarrow i + 1$
foreach $(x, y) \in ColourImage$ **do**
  if $SegmIm(x, y) = 1$ **then**
    $(X, Y, Z) \leftarrow CoordinateMapper(x, y)$
    $u \leftarrow W \frac{X - \min(XCoords)}{\max(XCoords) - \min(XCoords)}$
    $v \leftarrow H \frac{Y - \min(YCoords)}{\max(YCoords) - \min(YCoords)}$
    $ColourImage(x, y) \leftarrow \frac{1}{255} IrImage(x, y) \cdot TextureImage(u, v)$

Details about each part of the above process are presented in what follows. It should be noted that the computational complexity is linear, $O(n)$, in terms of the number of pixels of the reference texture, n, and it can be effectively run in real-time.

2.1. Segmentation

In the context of the IRT, the segmentation of a shirt from an image is not the main focus, as there are several available techniques that can be readily utilised [29,30,28]. We first use the depth information from the Kinect II to effectively remove the background. It returns stable results for various test cases used for validation. The output should be further refined in order to obtain the exact area corresponding to the garment that is later going to be retextured. In simple cases, handling the latter task would be possible through thresholding based on the colour. However, the aim of this paper is to create a retexturing algorithm that would suffice for garments with different colours, which might also contain a complicated texture, under different poses of the subject, with an arbitrary background setup. It is also desired to take the cases where different parts of the garment, e.g. shirt and pants, have a similar colour, into consideration, as well as the ones where the garment has a colour that is too close to the skin colour.

We try out two different approaches for refining the foreground, i.e. removing the parts of the background-removed image which do not belong to the garment. As the first solution, the prior information about which part of the previously segmented out person corresponds to the garment that needs to be retextured is incorporated into the Breadth-First Search (BFS) algorithm. The foregoing information can be obtained from the user as an interactive input, where they specify boundary lines at the edges of the restricted areas, which is necessary to make the algorithm avoid including parts of the person that should not be associated with the garment in the final segmentation result. The BFS possesses linear complexity. On the colour image, it starts from a prescribed point, and then expands to each pixel in the 4-neighbourhood of the part explored, unless the neighbourhood pixel is marked as restricted. Finally, the algorithm adds all the pixels that it has been able to reach to the final segmented garment.

The second approach, which is incorporated into the IRT in order to ensure that the proposed retexturing algorithm requires the minimal amount of manual contribution from the user, i.e. is as automatic as possible, is the interactive, semi-automatic method GrabCut [28], where the user has to leave marks, i.e. points or curves, that could help the software to distinguish the parts that are going to be left out from those expected to be kept. However, still the proportion of the amount of interaction with the user to the whole workload depends on a variety of factors, including the complexity of the scene, e.g. the type, intensity and direction of the illumination, the pose of the subject and the intricacy of the colours involved. For example, if the cloth is coloured similarly to the skin or the surrounding background, then a considerable level of input from the user is required, since in those cases, it is usually easier for the segmentation algorithm to confuse the garment with other ingredients of the scene, which means that a higher number of rounds of iterations receiving the guidance marks is needed to achieve the precision discriminating between the nuances involved.

On the other hand, if parts of the body or other components of the scene cover the garment, and appear on top of it in the image, which usually causes the garment to be divided into multiple zones, perhaps being too small, and not having regular shapes, much more interaction with the user is required for separating the parts of the garment that have to be retextured from the rest. The foregoing issues will be demonstrated more clearly, through the experimental results, in the next section. Nevertheless, typical scenarios could usually be dealt with in a reliably fast manner.

2.2. Texture mapping

Mapping the reference texture to the input image, which is subject to retexturing, is performed automatically, where the location of the pixel on the reference texture to take the colour value from is determined based on the normalized depth value. It is worth noticing that the reference texture image is assumed to possess ideal lighting conditions, and expected not to require preprocessing. The Kinect API enables the translation of screen coordinates of an image into real world coordinates. More specifically, the following map is used:

$$\omega : (x, y) \rightarrow (X, Y, Z), \tag{1}$$

where $x, y \in \mathbb{N}$ denote the screen coordinates of the image, and X, $Y, Z \in \mathbb{R}$ stand for the real-world coordinates corresponding to them. The objective is to map the screen coordinates onto that of the texture domain. It is achieved using the functions f_u and f_v, as follows:

$$u = f_u(x, y) = W \, \frac{\omega_x(x, y) - X_{min}}{X_{max} - X_{min}}, \tag{2}$$

$$v = f_v(x, y) = H \, \frac{\omega_y(x, y) - Y_{min}}{Y_{max} - Y_{min}}, \tag{3}$$

where u and v are the new texture domain coordinates, $\omega_x(x,y)$ and $\omega_y(x,y)$ denote the X and Y coordinates, respectively, obtained from the map $\omega(x,y)$, and X_{min}, Y_{min}, X_{max} and Y_{max} stand for the smallest and largest X and Y coordinates of the pixels from the blob obtained in the last step, respectively. In addition, W and H are the width and height of the texture image, respectively.

It should be noted that the texture space coordinates that are calculated in the above manner do not depend on the real-world Z-coordinates, i.e. the map from (x,y) to (u,v) is defined such that it avoids the so-called perspective effect.

2.3. Shading

Now that the texture space coordinates for each pixel have been specified, it becomes necessary to calculate their new colour values. A standard approach to doing so would be to use the already existing colour information from the original image. However, this has a drawback of not being totally colour invariant, as darker colours have a much smaller amount of intensity changes. Instead, as the main contribution of the paper, a novel shading system is proposed, which uses the IR image of the scene to infer the lighting conditions, and takes them into account for the sake of realistic shading calculation. More clearly, for extracting the pure, unshaded colours, the IR values are used, as they directly indicate the amount of light on the surface. To find out the desired relationship between the pure colour values and the shaded ones, a linear map that transforms the IR image pixels to their counterparts on the RGB image is derived, and applied to the colour values taken from the reference texture image.

Let us assume that a colour $c=(r,g,b)$, with its red, green and blue components, r, g and b, respectively, is obtained from the texture image, and the colour of the corresponding pixel on the RGB image is $c'=(r',g',b')$, with a similar notation. Obviously, the colour c' depends on the scene lighting conditions, and the actual colour under perfect lighting conditions, where the whole surface is uniformly lit, is still unknown. Nevertheless, c' can be expressed as a linear multiplier of the original colour of the pixel, $c^*=(r^*,b^*,g^*)$, as follows:

$$c'=l(x,y)c^*=(lr^*,lb^*,lg^*), \tag{4}$$

where l denotes the light intensity at that point. Then the colour of the texture image can be treated as the new true colour in maximum lighting intensity, and l is used as a coefficient to transform it through the same map, and find the shaded colour, i.e. $lc=(lr,lg,lb)$.

In many cases, the values of (x,y) that are very close either to zero or one produce unrealistically dark or bright regions. To overcome this problem, in the context of the IRT, histogram stretching is used in order to confine its values within the range $\left[\frac{\alpha}{255},\frac{\beta}{255}\right]$, where the values of α and β have to be determined in a case-by-case manner, depending on the experimental setup, since they are under the influence of the lighting conditions. All the possible combinations of the foregoing parameters are checked, and the best configuration is selected based on visual inspection, opting for the most realistic results achieved through retexturing. Nevertheless, the process of determining these values is only required to be performed once at the training stage under a certain experimental setup, and then just used as they are when testing. For example, Fig. 2 shows sample retextured garment images with β fixed at 160 and varying α. As it can be seen from the foregoing image, lower values of α have led to more realistic, hence desirable, results.

So the lighting intensity approximation can be expressed as follows:

$$l(x,y)=\alpha+\frac{\beta-\alpha}{Ir_{max}-Ir_{min}}\cdot(Ir(x,y)-Ir_{min}), \tag{5}$$

where Ir_{max} and Ir_{min} are the maximum and the minimum IR intensities in the segmented image area, respectively.

In fact, the experiments have shown that the proposed shading method operates reliably under a wide range of conditions, as it is pattern-, colour- and pose-invariant, and the only necessary precondition is that the garment to be retextured be made of the same material as the one present in the original Kinect frame, so that, in order to avoid the effect of the albedo changes, the infrared reflection intensity would be approximately the same in different parts of the surface.

3. The related works

In recent years there has been an increase in research related to retexturing, partly because of the more widespread use of depth cameras like Microsoft Kinect [31]. There are different kinds of approaches:

1. Marker-based methods where the image has easily detectable reference points that allow it to accurately recover the surface based on the displacement of the markers;
2. Methods that are trying to construct an approximation of the surface without such prior knowledge and using general characteristics like image gradients;
3. Methods that use multiple viewpoints or depth cameras.

White and Forsyth describe a method [24] that uses prior knowledge of the texture of a surface and that allows them to choose markers. They are trying to match the markers with the ones on the transformed surface and to create a map between them. They are using a simplified approach where they have only a small number of colours, which enables them to guess the original colour on the image and to obtain lighting intensity at that point.

Pilet et al. [32] also perform image registration between an image in its original form with no rotations in uniform lighting conditions and the same image, which has different lighting, occlusions or rotations. They are using the Expectation Maximisation framework that also provides visibility and lighting maps. The resulting method is able to handle difficult lighting conditions and occlusions.

Guo et al. [33] describe a method that operates under the assumption that locally, the depth gradients are proportional to the intensity gradients. They compute the texture coordinates by solving an energy minimisation problem. The lighting problem is solved by converting to *YCbCr* colour space and using the Y component as the intensity value for the new texture colour.

Shen et al. describe a somewhat similar retexturing algorithm [34] that works on high dynamic range images. They use image luminance to create an approximate depth map, calculate gradients of the depth map, and compute new texture coordinates and colours based on the previously obtained depth gradient map.

Zelinka et al. describe a material replacement system [35] where the interactively chosen area is retextured by first approximating the surface normals by assuming the Lambertian reflection model and improving it with the Gaussian mixture model. The new texture is synthesised with the jump-map technique.

Hilsmann and Eisert describe a method [36] that assumes a uniformly coloured surface with a small, easily detectable textured area on it. The texture is tracked with a video sequence using optical flow, and the occlusions are estimated based on that motion. As the main original colour of the surface is previously known, it is possible to easily estimate the lighting. In the textured area the lighting is estimated by interpolation using the light intensity of the uniformly coloured area.

Possible approaches to perform image registration for transformed non-rigid surfaces are discussed by Bartoli and Zisserman [37]. They are using Radial Basis Functions for regularising the optical flow field. Chui and Rangarajan [38] also discuss the same problem and propose a general point matching framework for non-rigid surfaces that is based on thin-plate splines.

Pizarro and Bartoli [39] focused on solving the self-occlusion problem; they used marker-based templates to reconstruct a deformed surface, which enabled them to achieve good results in retexturing.

The self-occlusion problem can also be solved by multiple viewpoints from several cameras. This kind of stereo vision systems enables more accurate geometry reconstruction. This has been studied by Pritchard and Heidrich [40] and by Scholz and Magnor [41].

Hilsmann and Eisert [42] created a database of different body positions at different levels of details to allow interactive visualisation and retexturing of clothes. The models of their database are created by using multiple images from different recalibrated viewpoints. The retexturing process uses the details provided by these models.

Khan et al. [43] create a rough reconstruction of a 3D surface by assuming that 3D distances can be approximated using image intensity. Texture coordinates are approximated based on the obtained gradients of the constructed surface, and the shading is based on interpolating between the original colour and the new texture colour.

Guo et al. [44] approximate a 3D surface by creating a triangulation based on several gradient-based feature points of the image, and the approximation is further enhanced by employing a refinement process based on the Poisson equation. The shading is accomplished by combining the *YCbCr* luminance component of the original image with the chroma components from the texture image. This kind of approach works fairly well, but is unfortunately limited to regions with sufficiently high luminance components. In the case of darker surfaces, the *Y* component values are often very low and the resulting retextured region would also be very dark.

A somewhat similar approach is taken by Shen et al. [45], but they use gradients of image intensity instead of a triangulation.

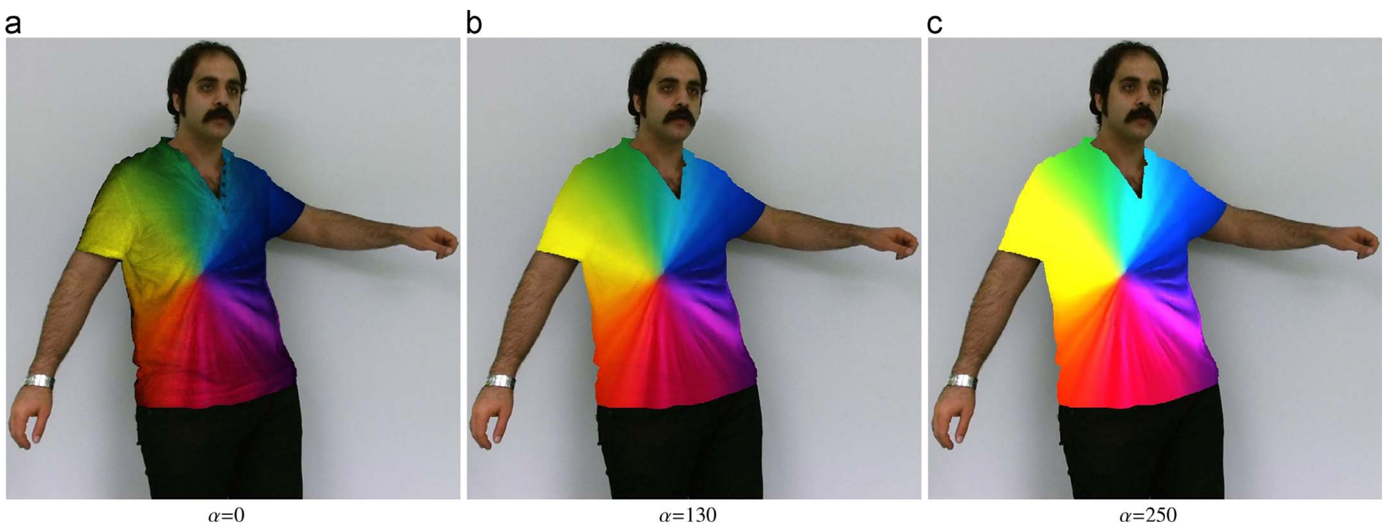

Fig. 2. The retexturing results obtained by fixing β at 160 and changing α. (a) α=0, (b) α = 130, (c) α=250.

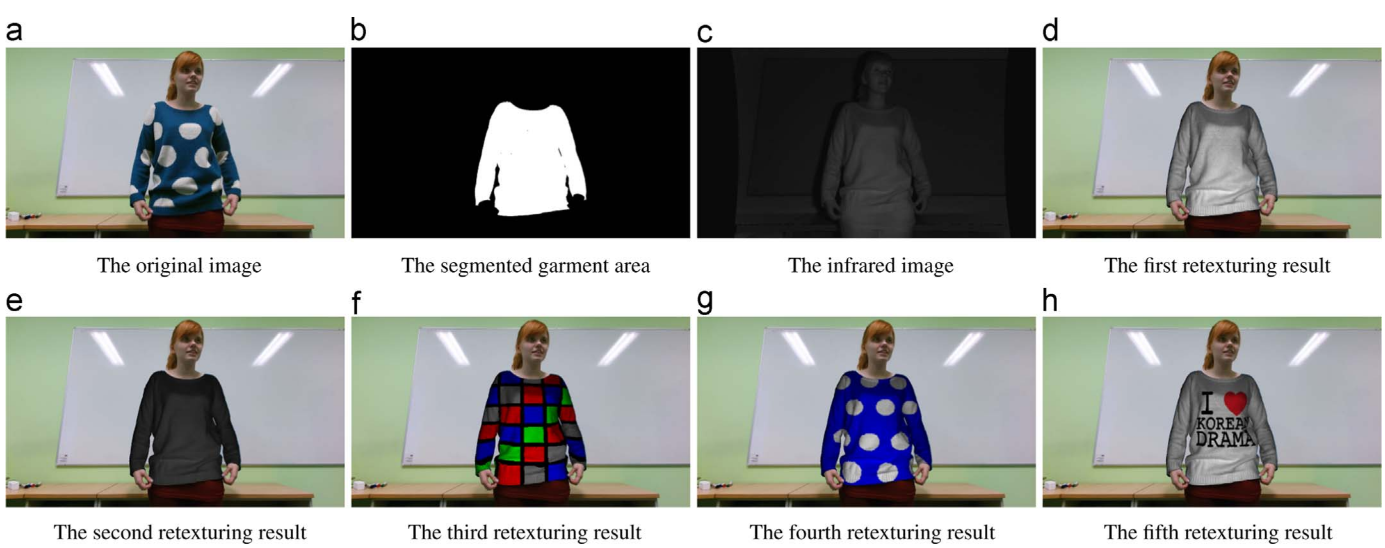

Fig. 3. The first example of the frames taken by the Kinect II camera, along with the results of retexturing by the IRT, with five different texture images. For segmentation, the BFS algorithm is used.

Based on the gradient field the Poisson equation is solved and the results are used to compute texture domain coordinates. The shading procedure is also based on the *YCbCr* colour space chroma replacement.

Kerl et al. [46] describe a simple method for estimating the pure albedo of the texture, in order to remove illumination effects from IR and colour images using Kinect II RGB-D sensors. They estimate the IR albedo from infrared and depth images, and transfer the former to the colour image. Using this approach, it is possible to create a colour shading model which includes all illumination effects and the colour albedo image.

In comparison with a relevant approach, Hongzhi Wu and Kun Zhou [47] introduce an interactive material acquisition system to capture the spatially varying appearance of daily objects, using Kinect. The system targets nonprofessional users, for capturing the appearance approximation of an object. The output is a 3D model that can be retextured and shaded. The object is scanned by the user who moves the Kinect sensor around the object. Our method has the advantage of retexturing an object from a single Kinect frame thereby reducing the errors caused by the movement of the object.

4. Results and discussion

First of all, it should be noted that the Kinect has to be properly calibrated, which could be accomplished using the process suggested in [48], making use of the OpenCV camera calibration [49] and the Kinect inverse disparity model [50]. In addition, the mismatch between the RGB and IR images could be alleviated by means of combining structured light with IR stereo [51], followed by analysis on the distribution of IR emitters [52], if applicable.

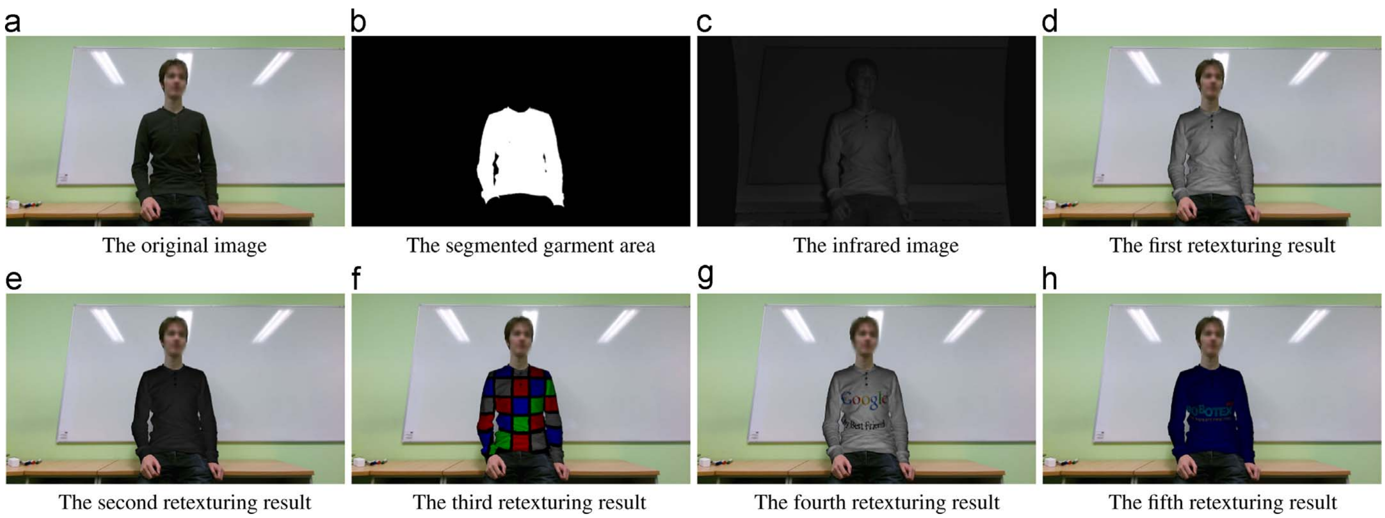

Fig. 4. The second example of the frames taken by the Kinect II camera, along with the results of retexturing by the IRT, with five different texture images. For segmentation, the BFS algorithm is used.

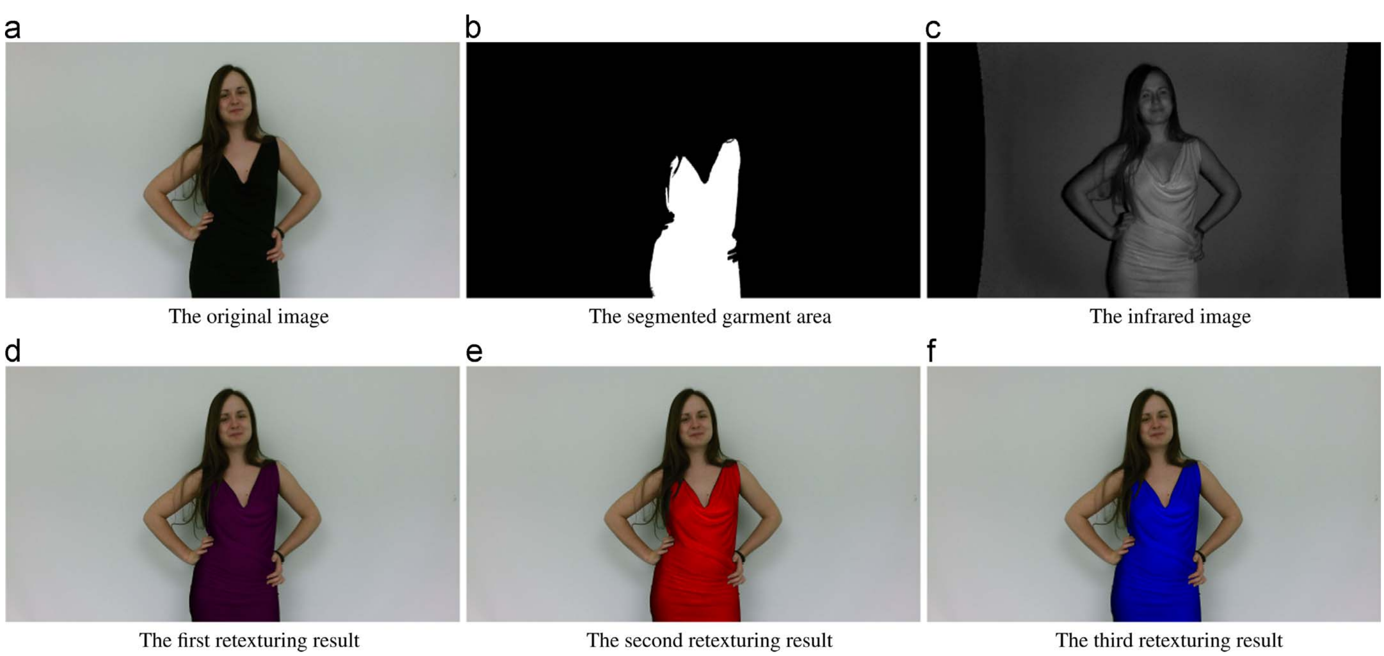

Fig. 5. The third example of the frames taken by the Kinect II camera, along with the results of retexturing by the IRT, with three different texture images. For segmentation, the BFS algorithm is used. This example shows one of the cases where due to the fact that the hair of the person appears on top of the garment, it is divided into multiple zones of different sizes and shapes, which incurs a higher amount of manual contribution needed from the user, to help the GrabCut algorithm to properly detect the areas that have to be retextured.

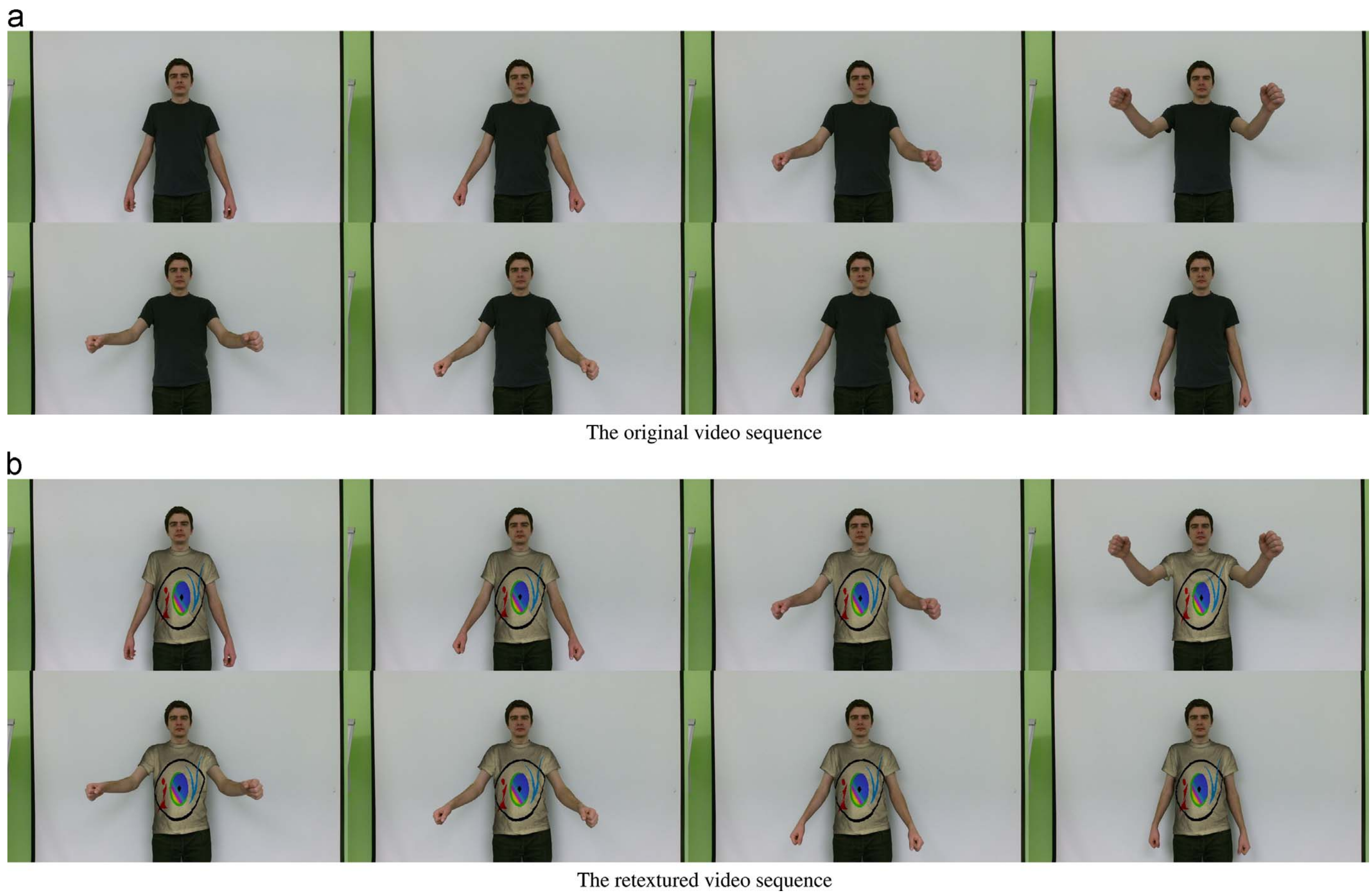

Fig. 6. Experimental results of implementing the proposed method on a video sequence. For segmentation, the GrabCut method is used.

Fig. 7. Experimental results demonstrating the robustness of the proposed method under variations of the lighting conditions, while introducing light sources of different types and directions, including ceiling and study lights. The second row shows the retextured counterparts of the images in the first row. From left to right, the images are presented from the darkest to the brightest ones. For segmentation, the GrabCut method is used.

The experimental results of applying the IRT, using the BFS algorithm for segmentation, with different reference textures, along with the original frames, are shown in Figs. 3–5, which show that the IRT produces realistic representations.

It is worth noticing that Fig. 5 demonstrates one of the cases where the amount of manual input from the user will be higher than average. The reason in most of those cases is that parts of the body or other components of the scene appear on top of the garment, and divide its image into different zones, meaning that the user will have to use the GrabCut more carefully, specifying a higher number of points and curves, that would help it to handle the boundaries of the smaller zones involved as well. For example, in the foregoing figure, the hair of the person appears on top of the garments, and divides its image into multiple parts, which might be comparatively smaller. Nevertheless, the retexturing results show that the IRT is capable of handling such cases properly.

Moreover, in order to demonstrate the efficiency of the proposed method while being applied to video sequences, sample implementation results are shown in Fig. 6, which could be more clearly illustrated through the video available online, at http://tinyurl.com/zwxlqrp. In addition, for examining its robustness against possible changes in the lighting conditions, including direction and intensity, as well as the albedo variations of the cloth, and to ensure that the underlying notion, i.e. transferring the IR information calculated on the Kinect II camera to that of an indoor environment, is sound, the images taken under varying lighting and environmental conditions are tested, and the results are shown in Fig. 7, where light sources of different types, including ceiling and study lights, are introduced from different directions, but have not affected the performance of the algorithm. Furthermore, for making sure that subject poses do not make negative effects on the performance of the IRT, it is applied to images of three subjects containing such variations, using the textures shown in Fig. 9. The sample images are shown in Fig. 8, along with their retextured counterparts, before and after shading, being intended to demonstrate the effect and necessity of the latter process in terms of realistic representation of all the details, such as wrinkles. The results of retexturing the garments on the same subjects while taking different poses, using the IRT, are illustrated in Fig. 10. The foregoing figures clearly substantiate the robustness of the proposed retexturing system against numerous changes of circumstances caused by the alterations of the real-life physical conditions that might affect the practical use, although they are not directly addressed in this study.

As most of the state-of-the-art methods rely on marker-based or motion-based surface reconstruction, there are not many that are suitable for comparison. On the other hand, all of the markerless approaches depend on the colour of the image, hence making the comparison meaningless when using the textured garment we are capable of dealing with. The most suitable choice for comparison would be a method that is also using a depth camera. Also, by using the infrared info we have a major advantage that enables us to achieve colour and texture invariance that none of the state-of-the-art methods have. It would be illogical to compare the results using images of a textured garment for methods that are not designed to handle it. For that reason, we use a constrained set of reference textures for comparison: the reference texture shows garment with no texture, but some garments have dark colours to demonstrate the inability of the state-of-the-art methods to handle more challenging conditions and to demonstrate the generality of our approach.

The results are compared with the method proposed by Shen et al. [45] and with the method of Khan et al. [43]. These methods were selected for comparison as they operate under conditions somewhat similar to our method: Both methods use static images, and do not depend on any prior knowledge about the surface pattern i.e. they do not use any prepared markers for surface reconstruction. For these reasons the method of Shen, Sun and the method of Khan, Reinhard represent the state-of-the-art methods that are the most suitable choices for comparison.

For the sake of clarity about the conditions of the comparison we feel that it is necessary to mention that both the methods of Shen et al. [45] and Khan et al. [43] use several input parameters that need to be carefully adjusted to achieve results as good as presented in their respective papers. In our comparison we experimentally fixed those parameters with sensible values that seemed to work reasonably well for most reference textures, but we did not do any additional parameter tuning, as we are

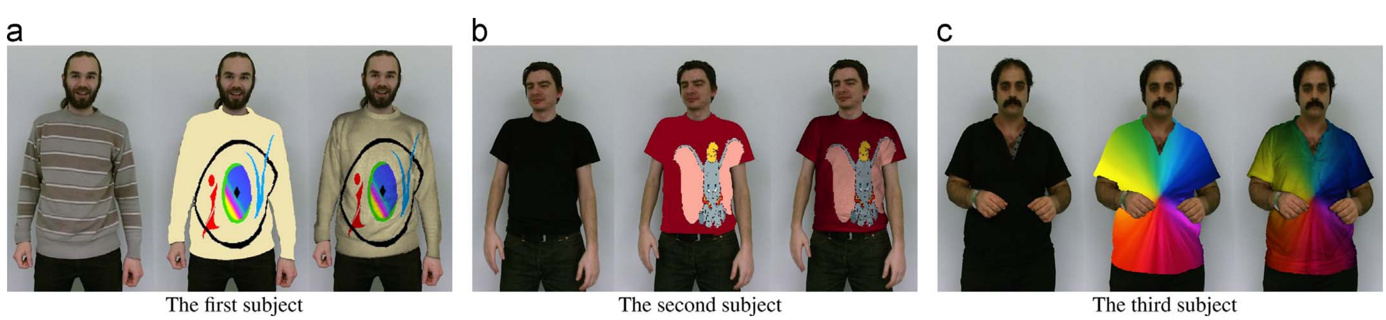

Fig. 8. Each sub-figure, from left to right, shows one of the subjects considered for testing the IRT on varying poses, with the original clothing and a sample retextured image, before and after shading, respectively. For segmentation, GrabCut is used.

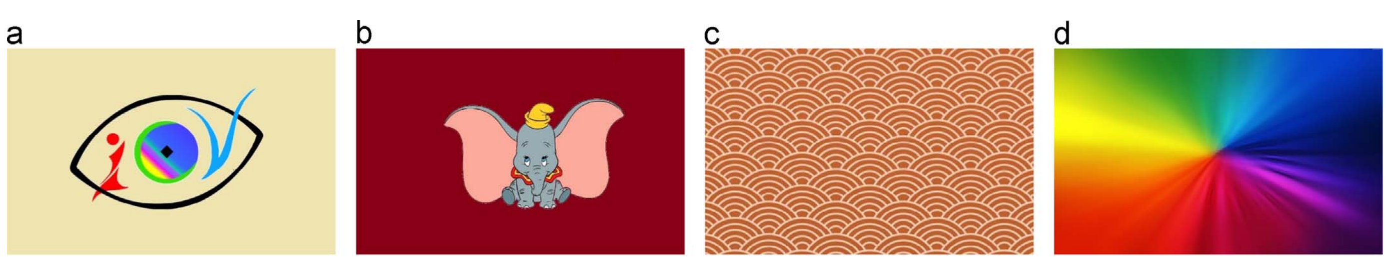

Fig. 9. The texture images used for conducting the experiments using GrabCut.

interested in comparing results that can be achieved automatically and not results that could be achieved when using them as interactive methods.

As there is no good metric to use for comparison, the mean opinion score (MOS) was used. A number of retextured images created by the IRT method, by the method of Shen and Sun [45], and by the method of Khan and Reinhard [43] were shown to a group of 40 people and everyone was asked to give their opinion on which of the displayed three images looked the most realistic.

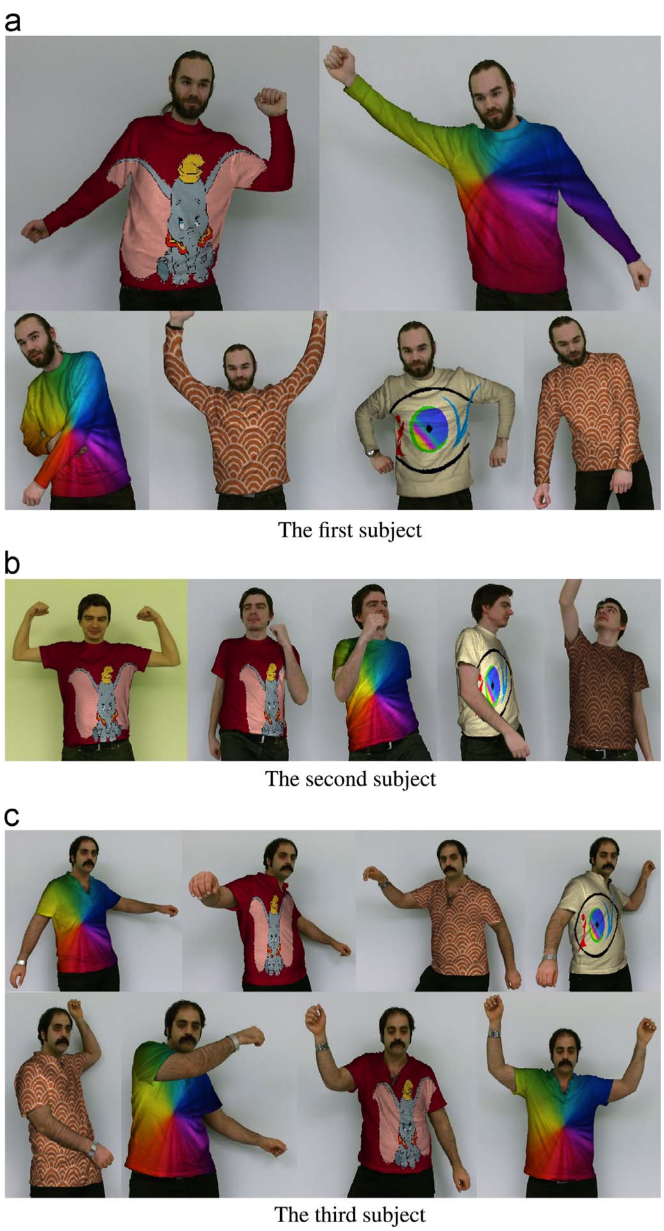

Fig. 10. The results of the implementation of the proposed method on instances of the subjects shown in Fig. 3 with varying poses. For segmentation, the GrabCut method is used.

Table 1
The mean opinion score (MOS) results.

Method	Mean opinion score (MOS)
IRT	566 votes
Method of Shen et al. [45]	57 votes
Method of Khan et al. [43]	177 votes

In total there were 20 images created by the IRT method, 20 images created by the method of Shen, Sun and 20 images created by the method of Khan, Reinhard. The overall results are shown in Table 1 and results for each individual image separately are shown in Fig. 11. The MOS results indicate that the IRT retexturing method produces more realistic retextured images than the state-of-the-art methods that were used for comparison. It can be seen that the state-of-the-art methods achieve results of similar quality in cases where the garment has brighter colours, but in case of darker garments the results are clearly inferior according to the MOS measurement.

A subset of the images used in MOS measurement are shown in Fig. 12. Images (a), (d), (g), (j) were created with the IRT retexturing method, images (b), (e), (h), (k) were created using the method of Shen, Sun [45] and images (c), (f), (i), (l) were created using the method of Khan et al. [43] using the same original image as input. In the case of the first 6 images (a)–(f), the images created by the IRT method were assessed to be much more realistic according to the MOS measurement. Images (g)–(l) demonstrate cases where the state-of-the-art techniques used for comparison achieved results of approximately the same quality as the IRT method according to the MOS measurements.

Inspecting the images, one can see that in images (b), (c), (e), (f) the lighting is quite unrealistic, whereas in images (a) and (d) the lighting looks much more natural. That is probably also the reason why images (a) and (c) achieved higher MOS scores. Images (g)–(l) have a somewhat more similar quality of surface lighting. In these cases the state-of-the-art methods were assessed to be almost as good as the IRT method.

Overall it is shown that the method of Shen, Sun and the method Khan, Reinhard both depend greatly on the original shirt colour, whereas the IRT method does not depend on the original colour or texture of the shirt. As expected, the other state-of-the-art methods fail to produce realistic results when the original colour of a shirt is dark and does not have much change in colour intensity. The IRT method on the other hand somewhat depends on the location of the IR emitter. If it is not located near the light source then it might cause some lighting inconsistency artifacts. Globally, the visual results and the MOS measurements indicate that the proposed IRT method performs better than the state-of-the-art methods in various conditions, producing much more stable results and realistic retexturing. Additionally the method is colour and texture invariant.

Fig. 5 illustrates the results of the IRT retexturing method on a dress. Overall, the aforementioned experimental results show that the generality of the IRT is sufficient for it to be applied to various types of garments. In principle it can create a realistic retexturing of any segmented object of the 3D scene.

The computational performance was also measured for the IRT method and the state-of-the-art comparison methods. All the time measurements were performed on a modern computer (Lenovo

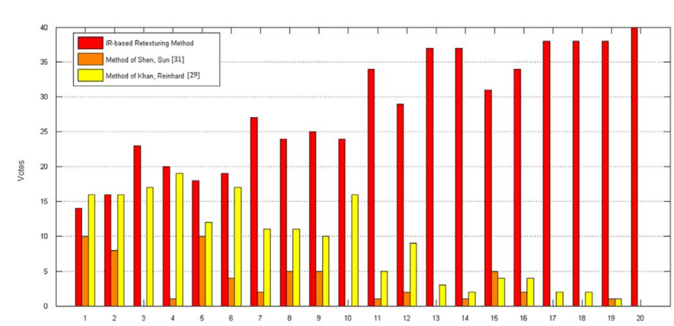

Fig. 11. MOS results for each individual image.

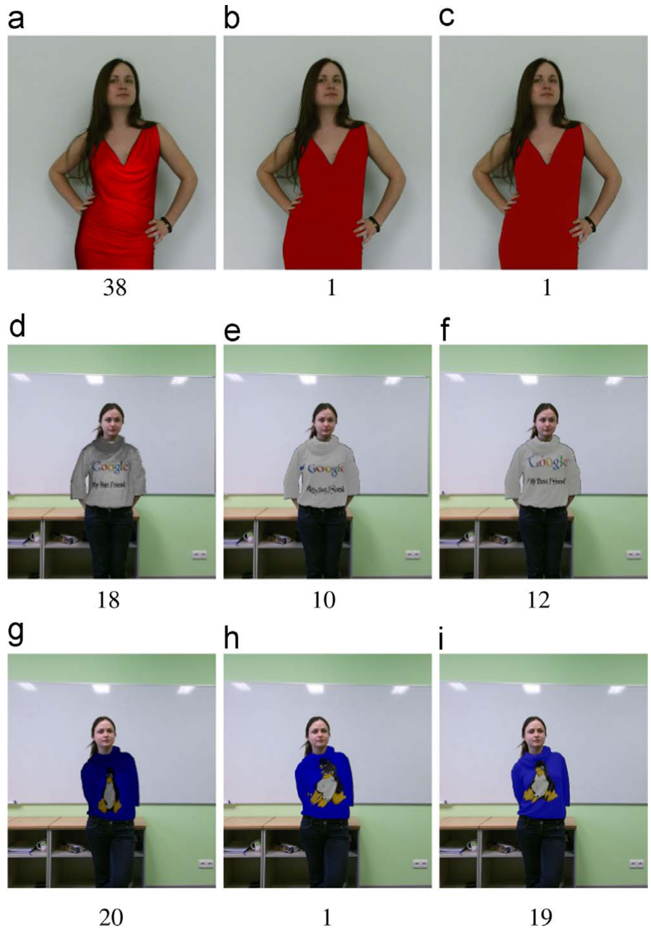

Fig. 12. A subset of the images that were used in the MOS comparison. In each row, from left to right, the images are created with the IRT, the method of Shen, Sun [45] and the method of Khan, Reinhard [43], respectively. The number of votes each image won is in its caption. For segmentation, the BFS algorithm has been used.

Table 2
Average computation times.

Method	Time (s)
IRT Method	0.463
Method of Shen et al. [45]	14.232
Method of Khan et al. [43]	0.297

Y510p) and the average of 20 measurements is shown in Table 2. It can be seen that the IRT method and the method of Khan et al. [43] can be used in real time scenarios, while the method of Shen et al. requires more computational time. This is mainly casued by the need to solve a very large linear system in order to solve the Poisson equation. It is likely that the performance of all the methods could be slightly improved by careful optimisation, but the change should not be significant, at least not without reducing the quality of the results.

It is worth noting that the notion of the IRT proposed in this paper is adjusted to the characteristics of single-image applications. Nevertheless, if extended properly, it can be used while incorporating the temporal information as well, for automatic segmentation of humans in video sequences, which should be paid due attention by the authors in the upcoming studies. The idea suggested in [53] could be adopted for this purpose, where the GrabCut is implemented only on the first frame, and the resulting Gaussian Mixture Models (GMM) are propagated throughout the next frames, thereby minimising the amount of input required from the user, and automatising the rest of the segmentation process.

5. Conclusion

This paper proposed a method for retexturing using the colour, depth and infrared information provided by the Microsoft Kinect II camera, with specific focus on shirt retexturing aimed at virtual fitting applications. The shirt was semi-automatically segmented out from the image, based on the colour markers and the depth information, using the GrabCut algorithm. Depth and corresponding 3D coordinates of the scene were used to create a texture map from the surface to the reference texture. As the main contribution of the paper, a novel approach was proposed for shading the textured surface using the infrared information, as it indicates the light intensity at different parts of the surface, according to which a linear transformation could be derived mapping the pure colours to that of the image, being then applied to the reference texture so as to mimic the illumination conditions in a realistic manner. The results were deemed superior to that of the state-of-the-art alternatives, as due to the foregoing novelty, the proposed system is illumination- and pose-invariant. Future works in this area should pay more particular attention to the problems of self-occlusions and image albedo variations. As regards the ongoing works, temporal information should be considered in the analyses, which was not yet addressed in this paper, due to the fact that it was focused only on single-image applications. For doing so, a simple solution is just to perform GrabCut semi-automatic segmentation on the first frame of a sequence, and transfer the Gaussian Mixture Models to the subsequent frames, so that the interaction for the segmentation task will be only required on the first frame, and posterior segmentation and retexturing will be fully automatic.

Acknowledgement

This work is supported Estonian Research Council Grant (PUT638) and the Spanish Project TIN2013-43478-P.

References

[1] Heckbert PS. Survey of texture mapping. IEEE Comput Graph Appl 1986;6(11):56–67.
[2] Maillot J, Yahia H, Verroust A. Interactive texture mapping. In: Proceedings of the 20th annual conference on computer graphics and interactive techniques. ACM, Anaheim, CA, USA; 1993, p. 27–34.
[3] Callieri M, Cignoni P, Corsini M, Scopigno R. Masked photo blending: mapping dense photographic data set on high-resolution sampled 3d models. Comput Graph 2008;32(4):464–73.
[4] Sun X, Xie G, Dong Y, Lin S, Xu W, Wang W, et al. Diffusion curve textures for resolution independent texture mapping. ACM Trans Graph 2012;31(4):74–82.
[5] Izadi S, Kim D, Hilliges O, Molyneaux D, Newcombe R, Kohli P, et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In: Proceedings of the 24th annual ACM symposium on user interface software and technology. ACM; 2011, p. 559–68.
[6] Sander PV, Snyder J, Gortler SJ, Hoppe H. Texture mapping progressive meshes. In: Proceedings of the 28th annual conference on computer graphics and interactive techniques. ACM, Los Angeles, CA, USA; 2001, p. 409–16.
[7] Bier EA, Sloan KR. Two-part texture mappings. IEEE Comput Graph Appl 1986;6(9):40–53.
[8] Hanrahan P, Haeberli P. Direct WYSIWYG painting and texturing on 3D shapes. In: ACM SIGGRAPH computer graphics, vol. 24. ACM; 1990, p. 215–23.
[9] Kim Y, Winnemoller H, Lee S. WYSIWIG stereo painting with usability enhancements. IEEE Trans Vis Comput Graph 2014;20(7):957–69.
[10] Xu S, Keyser J. Texture mapping for 3d painting using geodesic distance. In: Proceedings of the 18th meeting of the ACM SIGGRAPH symposium on interactive 3D graphics and games. ACM, San Francisco, CA, USA; 2014. p. 163.

[11] Traumann A, Daneshmand M, Escalera S, Anbarjafari G. Accurate 3d measurement using optical depth information. Electron Lett 2015;51(18):1420–2.
[12] Lawonn K, Gasteiger R, Rössl C, Preim B. Adaptive and robust curve smoothing on surface meshes. Comput Graph 2014;40:22–35.
[13] Kim D, Chae Y, Shin J, Baek U, Kim S. Interactive mirror system based on personal purchase information. Int J Softw Eng Appl 2013;7:3.
[14] Decaudin P, Julius D, Wither J, Boissieux L, Sheffer A, Cani MP. Virtual garments: a fully geometric approach for clothing design. In: Computer graphics forum, vol. 25. Wiley Online Library; 2006, p. 625–34.
[15] Brouet R, Sheffer A, Boissieux L, Cani MP. Design preserving garment transfer. ACM Trans Graph 2012;31(4):36.
[16] Chang WC, Chang WC. Real-time 3d rendering based on multiple cameras and point cloud. In: 2014 7th international conference on ubi-media computing and workshops (UMEDIA). IEEE; 2014. p. 121–6.
[17] Li P, Sun H, Huang C, Shen J, Nie Y. Interactive image/video retexturing using gpu parallelism. Comput Graph 2012;36(8):1048–59.
[18] Turquin E, Cani MP, Hughes JF. Sketching garments for virtual characters. In: ACM SIGGRAPH 2007 courses. ACM, San Diego, CA, USA; 2007. p. 28.
[19] Yasseen Z, Nasri A, Boukaram W, Volino P, Magnenat-Thalmann N, et al. Sketch-based garment design with quad meshes. Comput-Aid Des 2013.
[20] Zhang M, Lin L, Pan Z, Xiang N. Topology-independent 3d garment fitting for virtual clothing. Multimedia Tools Appl 2013:1–17.
[21] Zhou B, Chen X, Fu Q, Guo K, Tan P. Garment modeling from a single image. In: Computer graphics forum, vol. 32. Wiley Online Library; 2013. p. 85–91.
[22] Guo Y, Wang J, Sun H, Cui X, Peng Q. A novel constrained texture mapping method based on harmonic map. Comput Graph 2005;29(6):972–9.
[23] Gay-Bellile V, Bartoli A, Sayd P. Direct estimation of nonrigid registrations with image-based self-occlusion reasoning. IEEE Trans Pattern Anal Mach Intell 2010;32(1):87–104.
[24] White R, Forsyth D. Retexturing single views using texture and shading. In: Computer vision–ECCV 2006. Springer, Graz, Austria; 2006. p. 70–81.
[25] Watt AH, Watt A. 3D computer graphics, vol. 2. Reading: Addison-Wesley; 2000.
[26] Daneshmand M, Aabloo A, Ozcinar C, Anbarjafari G. Real-time automatic shape-changing robot adjustment and gender classification. Signal Image Video Process 2015:1–8.
[27] Bader DA, Madduri K. Designing multithreaded algorithms for breadth-first search and st-connectivity on the cray mta-2. In: International conference on parallel processing, 2006, ICPP 2006. IEEE; 2006. p. 523–30.
[28] Rother C, Kolmogorov V, Blake A. Grabcut: interactive foreground extraction using iterated graph cuts. ACM Trans Graph 2004;23(3):309–14.
[29] Fuhrmann A, Groß C, Luckas V, Weber A. Interaction-free dressing of virtual humans. Comput Graph 2003;27(1):71–82.
[30] Lee Y, Ma J, Choi S. Automatic pose-independent 3d garment fitting. Comput Graph 2013;37(7):911–22.
[31] Traumann A, Anbarjafari G, Escalera S. A new retexturing method for virtual fitting room using kinect 2 camera. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops; 2015. p. 75–9.
[32] Pilet J, Lepetit V, Fua P. Retexturing in the presence of complex illumination and occlusions. In: Proceedings of the 2007 6th IEEE and ACM international symposium on mixed and augmented reality. IEEE Computer Society, Nara, Japan; 2007. p. 1–8.
[33] Guo Y, Wang J, Zeng X, Xie Z, Sun H, Peng Q. Image and video retexturing. Comput Anim Virtual Worlds 2005;16(3–4):451–61.
[34] Shen J, Jin X, Sun H. High-dynamic range image tone mapping and retexturing using fast trilateral filtering. Vis Comput 2007;23(9–11):641–50.
[35] Zelinka S, Fang H, Garland M, Hart JC. Interactive material replacement in photographs. In: Proceedings of graphics interface 2005. Canadian Human–Computer Communications Society, Waterloo, Ontario, Canada; 2005. p. 227–32.
[36] Hilsmann A, Eisert P. Tracking and retexturing cloth for real-time virtual clothing applications. In: Computer vision/computer graphics collaboration techniques. Springer, Rocquencourt, France; 2009. p. 94–105.
[37] Bartoli A, Zisserman A, et al. Direct estimation of non-rigid registrations. In: British machine vision conference; 2004. p. 899–908.
[38] Chui H, Rangarajan A. A new point matching algorithm for non-rigid registration. Comput Vis Image Understand 2003;89(2):114–41.
[39] Pizarro D, Bartoli A. Feature-based deformable surface detection with self-occlusion reasoning. Int J Comput Vis 2012;97(1):54–70.
[40] Pritchard D, Heidrich W. Cloth motion capture. In: Computer graphics forum, vol. 22. Wiley Online Library; 2003. p. 263–71.
[41] Scholz V, Magnor MA. Cloth motion from optical flow. In: VMV, vol. 4; 2004. p. 117–24.
[42] Hilsmann A, Eisert P. Image-based animation of clothes. In: Eurographics (short papers); 2012. p. 69–72.
[43] Khan EA, Reinhard E, Fleming RW, Bülthoff HH. Image-based material editing. ACM Trans Graph 2006;25(3):654–63.
[44] Guo Y, Sun H, Peng Q, Jiang Z. Mesh-guided optimized retexturing for image and video. IEEE Trans Vis Comput Graph 2008;14(2):426–39.
[45] Shen J, Sun H, Mao X, Guo Y, Jin X. Color-mood-aware clothing re-texturing. In: 2011 12th international conference on computer-aided design and computer graphics (CAD/Graphics). IEEE, Jinan, China; 2011. p. 151–4.
[46] Kerl C, Souiai M, Sturm J, Cremers D. Towards Illumination-Invariant 3D Reconstruction Using ToF RGB-D Cameras. In: Second international conference on 3D vision (3DV), vol. 1. IEEE; 2014. p. 39–46.
[47] Wu H, Zhou K. AppFusion: Interactive Appearance Acquisition Using a Kinect Sensor. In: Computer Graphics Forum. Wiley Online Library; 2015.
[48] Burrus N. Kinect calibration. Available: 〈http://nicolasburrusname/indexphp/Research/KinectCalibration"** 〈http://nicolasburrusname/indexphp/Research/KinectCalibration〉2011;.
[49] Dryanovski I, Morris W, Magnenat S. Kinect node 2010.
[50] Garage W. Camera calibration and 3d reconstruction 2011.
[51] Alhwarin F, Ferrein A, Scholl I. Ir stereo kinect: improving depth images by combining structured light with ir stereo. In: PRICAI 2014: trends in artificial intelligence. Springer, Gold Coast, QLD, Australia; 2014. p. 409–21.
[52] Khoshelham K, Elberink SO. Accuracy and resolution of kinect depth data for indoor mapping applications. Sensors 2012;12(2):1437–54.
[53] Hernández-Vela A, Reyes M, Ponce V, Escalera S. Grabcut-based human segmentation in video sequences. Sensors 2012;12(11):15376–93.