FLORE
Repository istituzionale dell'Università degli Studi di Firenze

From 2D to 2.5D i.e. from painting to tactile model

Questa è la Versione finale referata (Post print/Accepted manuscript) della seguente pubblicazione:

Original Citation:

From 2D to 2.5D i.e. from painting to tactile model / Furferi R., Governi L., Volpe Y., Puggelli L., Vanni N., Carfagni M.. - In: GRAPHICAL MODELS. - ISSN 1524-0703. - STAMPA. - Volume 76, Issue 6:(2014), pp. 706-723. [10.1016/j.gmod.2014.10.001]

Availability:

The webpage https://hdl.handle.net/2158/904939 of the repository was last updated on 2016-09-01T15:19:36Z

Published version:

DOI: 10.1016/j.gmod.2014.10.001

Terms of use:

Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf)

Publisher copyright claim:

La data sopra indicata si riferisce all'ultimo aggiornamento della scheda del Repository FloRe - The above-mentioned date refers to the last update of the record in the Institutional Repository FloRe

(Article begins on next page)

09 November 2026

Graphical Models 76 (2014) 706–723

Contents lists available at ScienceDirect

Graphical Models

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

From 2D to 2.5D i.e. from painting to tactile model ☆

Rocco Furferi *, Lapo Governi, Yary Volpe, Luca Puggelli, Niccolò Vanni, Monica Carfagni

Department of Industrial Engineering of Florence, Via di Santa Marta 3, 50139 Firenze, Italy

ARTICLE INFO

Article history:
Received 9 August 2014
Received in revised form 10 October 2014
Accepted 13 October 2014
Available online 22 October 2014

Keywords:
2.5D model
Shape From Shading
Tactile model
Minimization techniques

ABSTRACT

Commonly used to produce the visual effect of full 3D scene on reduced depth supports, bas relief can be successfully employed to help blind people to access inherently bi-dimensional works of art. Despite a number of methods have been proposed dealing with the issue of recovering 3D or 2.5D surfaces from single images, only a few of them explicitly address the recovery problem from paintings and, more specifically, the needs of visually impaired and blind people.

The main aim of the present paper is to provide a systematic method for the semi-automatic generation of 2.5D models from paintings. Consequently, a number of *ad hoc* procedures are used to solve most of the typical problems arising when dealing with artistic representation of a scene. Feedbacks provided by a panel of end-users demonstrated the effectiveness of the method in providing models reproducing, using a tactile language, works of art otherwise completely inaccessible.

© 2014 Elsevier Inc. All rights reserved.

1. Background

Haptic exploration is the primary action that visually impaired people execute in order to encode properties of surfaces and objects [1]. Such a process is based on a cognitive iter based on the combination of somatosensory perception of patterns on touched surface (e.g., edges, curvature, and texture) and proprioception of hand position and conformation [2]. As a consequence, visually impaired people's cognitive path is significantly restrained when dealing with experience of art. Not surprisingly, in order to confront with this issue, the access to three-dimensional copies of artworks has been the first degree of interaction for enhancing the visually impaired people experience of art in museums and, therefore, numerous initiatives based on the interaction with sculptures and tactile three-dimensional reproductions or architectural aids on scale have been developed all around the world. Unfortunately, cultural heritage in the form of two-dimensional art (e.g. paintings or photographs) are mostly inaccessible to visually impaired people since they cannot be directly reproduced into a 3D model. As of today, even if numerous efforts in translating such bi-dimensional forms of art into 3D model are in literature, sill a few works can be documented. With the aim of enhancing experience of 2D images, painting subjects have to be translated into an appropriate "object" to be touched. This implies a series of simplifications in the artworks "translation" to be performed according both to explicit users suggestions and to scientific findings of the last decades.

In almost all the scientific work dealing with this subject, a common objective is shared: to find a method for translating paintings into a simplified (but representative) model meant to help the visually impaired people in understanding both the painted scene (position in space of painted subjects) and the "shape" of subjects themselves. On the basis of recent literature, a wide range of simplified models can be built to provide visually impaired people with a faithful description of paintings [3]. Among

☆ This paper has been recommended for acceptance by Ralph Martin and Peter Lindstrom.

* Corresponding author. Fax: +39 (0)554796400.
E-mail address: rocco.furferi@unifi.it (R. Furferi).

http://dx.doi.org/10.1016/j.gmod.2014.10.001
1524-0703/© 2014 Elsevier Inc. All rights reserved.

them, the following representations proved to be quite effective, as experimentally demonstrated in [4]: tactile diagrams (e.g. tactile outline-based and texturized pattern-based reconstruction) and bas-relief (e.g. flat layered bas-relief, shaped bas-relief).

Tactile diagrams are not a relief reproduction of visual images: rather, they are translations of visual images into a tactile language consisting on the main outlines of the subjects to be explored mixed together with patterns added to discriminate different surfaces characterizing the scene and/or the position of painted subjects. The most common way for creating such a representation is to separate the background and foreground, or the ground and figures, illustrating them in two separate diagrams or using two different patterns. More in general, outline-based representations may be enriched using different textures each one characterizing different position in space and/or different surface properties [4]. Moreover, these representations are generally used in conjunction with verbal narratives that guides the user through the diagram in a logical and ordered manner [5,6].

Unlike tactile diagrams, bas-relief representation of paintings delivers 3D information in a more "realistic" way by improving depth perception. For this reason, as demonstrated in the cited recent study [4], this kind of model (also referred as "2.5D model"), is one of the most "readable" and meaningful for blind and visually impaired people; it proves to provide a better perception of painted subjects shape and, at the same time, a clear picture of their position in the scene. Moreover, according to [7], the bas-relief is perceived as being more faithful to the original artwork and the different relief (height) assigned to objects ideally standing on different planes is considered very useful in discriminating foreground objects from middle ground and background ones.

The topic of bas-relief reconstruction starting from a single image is a long-standing issue in computer vision scientific literature and moves from the studies aiming at recovering 3D information from 2D pictures or photographs. Two of the best known works dealing with this topic are [8,9]. In [8] a method for extracting a non-strictly three-dimensional model of a painted scene with single point perspective is proposed, making use of vanishing points identification, foreground from background segmentation and polygonal reconstruction of the scene. In [9] a coarse, scaled 3D model from a single image by classifying each image pixel as ground, vertical or sky and estimating the horizon position is automatically built in the form of a pop-up model. These impressive works, as well as similar approaches, however, are not meant to create a bas-relief; rather they are aimed to create a 3D virtual model of the scene when objects are virtually separated each other.

Explicitly dealing with relief reconstruction from single images, some significant studies can be found in literature, especially dealing with coinage and commemorative medals (see for instance [10,11]). In most of the proposed approaches the input image, often representing human faces, stemmas, logos and figures standing out from the image background are translated into a flat bas-relief by adding volume to the represented subjects. The easier, and probably best known, method to perform a bas-relief reconstruction from images is image embossing [12,13], a widely recognized computer graphics technique in which each pixel of an image is replaced either by a highlight or a shadow, depending on boundaries on the original image. The result obtained using this technique consists of a relief visually resembling the original image but affected by shallow and incorrect depth reconstruction (due to the algorithm based on image gradient computation). Some improvements to this method, like the one proposed in [14], just to cite a few, are in scientific literature; the embossing method is enhanced by using pre-processing techniques based on image enhancement, histogram equalization and dynamic range. The uses of unsharp masks and smoothing filters have also been extensively adopted to emphasize salient features and deemphasize others in the original image, so that the final results better resemble the original image. In a recent paper [15] an approach for estimating the height map from single images representing brick and stone reliefs (BSR) has been also proposed. The method proved to be adequate for restoring BSR surfaces by using a height map estimation scheme consisting of two levels: the bas-relief, referring to the low frequency component of the BSR surfaces, and the high frequency detail. Commercial software, like ArtCAM and JDPaint [16] have been also developed making available functions for bas-relief reconstruction from images. In these software packages users are required to use vector representation of the object to be reconstructed and "inflate" the surface delimited by the object outlines. The above cited method prove to be effective in creating models where the subjects are volumetrically detached from the background but with compressed depth [3] like, for instance, models resembling figures obtained embossing a metallic plate. In order to obtain a faithful surface reconstruction a strong interaction is required; in particular for complex shapes, such as faces, it is not sufficient to vectorialize the subject's outlines but each part to be inflated needs to be outlined and vectorialized. In case of faces, for example, lips, cheeks, nose, eyes, eyebrows etc. shall be manually drafted. This is a time-consuming task when dealing with paintings, often characterized by a number of subjects blended into the background (or by a background drawing attention from the main subjects).

The inverse problem of creating 2.5D models starting from 3D models has also been extensively studied [17,18]. These techniques use normal maps obtained by the (available) 3D model and uses techniques for performing the compression in the 2.5D space. However, these works are not suitable for handling the reconstruction from a 2D scene where normal map is the desired result and not a starting point.

On the basis of the above mentioned works, it is evident that the method researched in this paper is not fully explored and only a few works are in literature. One of the most important methods aimed to build models visually resembling sculptors-made bas-relief from paintings is the one proposed by [19]. In this work the high resolution image of the painting to be translated into bas-relief is manually processed in order to (1) extract the painted subject's contours, (2) to identify semantically meaningful

areas and (3) to assign appropriate height values to each area. One of the results consists of "a layered depth diagram" made of a number of individual shapes cut out of flexible sheets of constant thickness, which are glued on top of each other to resemble a painting". Once the layered depth is built, in the same work, a method for extracting textures from the painted scene and for assigning them to the 2.5D model is drawn. Some complex shapes, such as subject faces, are modeled using apposite software and the resulting normal maps are finally imported in the model and blended together. The result consists of a high quality texturized relief. While this approach may be extremely useful in case perspective related information is lacking in the considered paintings, it is not the best option in case of perspective paintings. In fact, the reconstructed scene could be geometrically inconsistent, thus causing misperception by blind people. This is mainly due to the fact that the method is required to manually assign height relations for each area regardless to the information coming from scene perspective. Moreover, using frequency analysis of image brightness to convey volumetric information can provide inconsistent shape reconstruction due to the concave–convex ambiguity. Nonetheless, the chief findings of this work are used as inspiration from the present work, together with other techniques aimed to retrieve perspective information and subjects/objects shapes, as described in the next sections. With the aim of reducing the error in reconstructing shapes from image shading, the most studied method is the so called Shape From Shading (SFS). Extensively studied in the last decades [20–23], SFS is a computational approach that bundles a number of techniques aimed to reconstruct the three-dimensional shape of a surface shown in a single gray-level image. However, automatic shape retrieval using SFS techniques proves to be unsuitable for producing high-quality bas-reliefs [17]; as a consequence, more recent work by a number of researchers has shown that moderate user-interaction is highly effective increasing 2.5D models from a single view [20–23]. Moreover, [24] proposed a two-step procedure: the first step recovers high frequency details using SFS; the second step corrects low frequency errors using a user-driven editing tool. However, this approach entails quite a considerable amount of user interaction especially in the case of complex geometries. Nevertheless, in case the amount of required user interaction can be maintained at a reasonable level, interactive SFS methods may be considered among the best candidate techniques for generating good quality bas-reliefs starting from single images.

Unfortunately, since paintings are hand-made artworks, many aspects of the represented scene (such as silhouette and tones) are unavoidably not accurately reproduced in the image, thus leading to an even more complex task in performing reconstruction with respect to the analysis of synthetic or real-world images. In fact, image brightness and illumination in a painting are only an artistic reproduction of a (real or imagined) scene. To make things even worse, light direction is unknown in the most of the cases and a diffused light effect is often added by the artist to the scene. Furthermore, real (and represented) objects surfaces may have complex optical properties, far from being approximated with Lambertian surfaces [22]. These drawbacks have great impact on the reconstruction: any method used for retrieving volume information from paintings shall be able to retrieve 3D geometry on the basis of defective information. As a consequence any approach for solving such a complex SFS-based reconstruction shall require a number of additional simplifying assumptions and a worse outcome is always expected with reference to results obtained for synthetic and real-world images.

With these strong limitations in mind, the present work provides a valuable attempt in providing sufficiently plausible reconstruction of artistically reproduced shaded subjects/objects. In particular the main aim is to provide a systematic user-driven methodology for the semi-automatic generation of tactile 2.5D models to be explored by visually impaired people. The proposed methodology lays its foundations on the most promising existing techniques; nonetheless, due to the considerations made about painted images, a series of novel concepts are introduced in this paper: the combination of the spatial scene reconstruction and the volume definition using different volume-based contributions, the possibility of modeling scenes with unknown illumination and the possibility of modeling subjects whose shading is only approximately represented. The method does not claim to perform a perfect reconstruction of painted scenes. It is, rather, a process intended to provide a plausible reconstruction by making a series of reasoned assumptions aimed at solving a number of problems arising from 2.5D reconstruction starting from painted images. The devised methodology is supported by a user-driven graphical interface, designed with the intent of helping non-expert users (after a short training phase) in retrieving final surface of painted subjects (see Fig. 1).

For sake of clarity, the description of the tasks carried out to perform reconstruction will be described with reference to an exemplificative case study i.e. the reconstruction of "The Healing of the Cripple and the Raising of Tabitha" fresco by Masolino da Panicale (see Fig. 2). This masterpiece is a typical example of Italian Renaissance paintings characterized by single-point perspective.

2. Method

With the aim of providing a robust reconstruction of the scene and of the subjects/objects imagined by the artist in a painting, the systematic methodology relies on an interactive Computer-based modeling procedure that integrates:

(1) Preliminary image processing-based operation on the digital image of a painting; this step is mainly devoted to image distortion correction and segmentation of subjects in the scene.
(2) Perspective geometry-based scene reconstruction; mainly based on references [7,8,25], but modeling also oblique planes, this step allows the reconstruction of the painted scene i.e. the geometric arrangement of painted subjects and objects in a 2.5D virtual space using perspective-related information when available. The final result of this step is a virtual "flat-layered bas-relief".

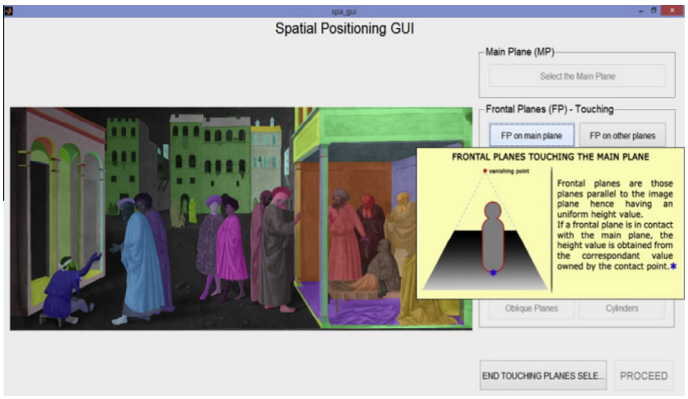

Fig. 1. A screenshot of the GUI, designed with the intent of helping non-expert users in retrieving final surface starting from painted images.

Fig. 2. Acquired image of "The Healing of the Cripple and the Raising of Tabitha" fresco by Masolino da Panicale in the Brancacci Chapel (Church of Santa Maria del Carmine in Florence, Italy).

(3) Volume reconstruction; using an appositely devised approach based on SFS and making use of some implementations provided in [26], this step allow the retrieval of volumetric information of painted subjects. As already stated, this phase is applied on painted subjects characterized by incorrect shading (guessed by the artist); as a consequence, the proposed method introduces, in authors' opinion, some innovative contributions.
(4) Virtual bas-relief reconstruction; by integrating the results obtained in steps 2–3, the final (virtual) bas-relief resembling the original painted scene is retrieved.
(5) Rapid prototyping of the virtual bas-relief.

2.1. Preliminary image processing-based operation on the digital image

A digital copy of the original image to be reconstructed in the form of bas-relief is acquired using proper image acquisition device and illumination. Generally speaking, image acquisition should be performed with the intent of obtaining a high resolution image which preserve shading since this information is to be used for virtual model reconstruction. This can be carried out using calibrated or uncalibrated cameras. Referring to the case study used for explaining the overall methodology, the acquisition device consists of a Canon EOS 6D camera (provided with a 36×24 mm^2 CMOS sensor with a resolution equal to 5472×3648 pixel^2). A CIE standard illuminant D65 lamp placed frontally to the painting was chosen in order to perform a controlled acquisition. The acquired image I_a (see for instance Fig. 2) is properly undistorted (by evaluating lens distortion) and rectified using widely recognized registration algorithms [27]. Let, accordingly, I_r (size $n \times m$) be the rectified and undistorted digital image representing the painted scene. Since both scene reconstruction and volumes definition are based on gray-scale information, the color of such an image has to be discarded. In this work, color is discarded by performing a color conversion of the original image from sRGB to CIE-LAB color space and by extracting the L^* channel. In detail, a first color conversion from sRGB color space to the tristimulus values CIE XYZ is carried out using the equations available for the D65 illuminant [28]. Then, the color transformation from CIE XYZ to CIELAB space is performed simply using the XYZ to CIELAB relations [28]. The final result consists of a new image $I_L^* a^* b^*$. Finally, from such an image the channel L^* is extracted thus defining a grayscale image Υ_L that can be considered the starting point for the entire devised methodology (see Fig. 3).

Once the image Υ_L is obtained, the different objects represented in the scene, such as human figures, garments,

architectural elements, are properly identified. This task is widely recognized with the term "segmentation" and can be accomplished using any of the methods available in literature (e.g. [29]). In the present work segmentation is performed by means of the developed GUI, where the objects outlines are detected using the interactive livewire boundary extraction algorithm [30]. The result of segmentation consists of a new image C (size $n \times m$) where different regions (clusters C_i) are described by a different label L_i (see for instance Fig. 4 where the clusters are represented in false colors) with $i = 1 \ldots k$ and where k is the number of segmented objects (clusters).

Besides image segmentation, since the volume definition techniques detailed in next sections are mainly based on the analysis of the shading information provided by each pixel of the acquired image, another important operation to be performed on image Υ_L consists of albedo normalization. In detail it is necessary to normalize, the albedo ρ_i of every segment, which is pixel-by-pixel the amount of diffusively reflected light, to a constant value. For sake of simplicity, it has been chosen to normalize the albedo to 1, and this is obtained by dividing the gray channel of each segment by its actual albedo value:

$$\Upsilon_i = \frac{1}{\rho_i} \Upsilon_{L_i} \tag{1}$$

The final results of this preliminary image-processing based phase are: (1) a new image Υ representing the grayscale version of the acquired digital image with corrected albedo and (2) clusters C_i each one representing a segment of the original scene.

2.2. Perspective geometry-based scene reconstruction

Once the starting image has been segmented it is necessary to define the properties of its regions, in order to arrange them in a consistent 2.5D scene. This is due to the fact that the devised method for retrieving volumetric information (described in the next sections) requires the subject describing the scene to be geometrically and consistently placed in the space, but described in terms of flat regions.

In other words, a deliberate choice is made by the authors here: to model, in the first instance, each subject

Fig. 4. Example of a segmented image (different color represents different segment/region). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(i.e. each cluster C_i) as a planar region thus obtaining a virtual flat-layered bas-relief (visually resembling the original scene but with flat objects) where the relevant information is the perspective-based position of subjects in the virtual scene. Two reasons are supporting this choice: firstly, psychophysical studies performed with blind people documented in [4] demonstrated that flat-layered bas-relief representation of paintings are really useful for a first rough understanding of the painted scene. Secondly, for each subject (object) of the scene, the curve identified on the object by the projection along the viewing direction (normal to the painting plane) approximately lays on a plane if the object is limited in size along the projection direction i.e. when the shape of the reconstructed object has limited size along the viewer direction (see Fig. 5). Since the final aim of the present work is to obtain a bas-relief where objects slightly detach from the background, this assumption is considered valid for the purpose of the proposed method.

A number of methods for obtaining 3D models from perspective scenes are in literature, proving to be very effective for solving this issue (see for instance [19]). Most of them, and in particular the method provided in [8], can be successfully used to perform this task. Or, at most, combining the results obtained using [8] with the method described in [19] for creating layered depth diagrams, this kind of spatial reconstruction can be accomplished. As stated in the introduction, however, one of the aims of the present work is to describe the devised user-guided

Fig. 3. Grayscale image Υ_L obtained using L^* channel of $I_{L^*a^*b^*}$.

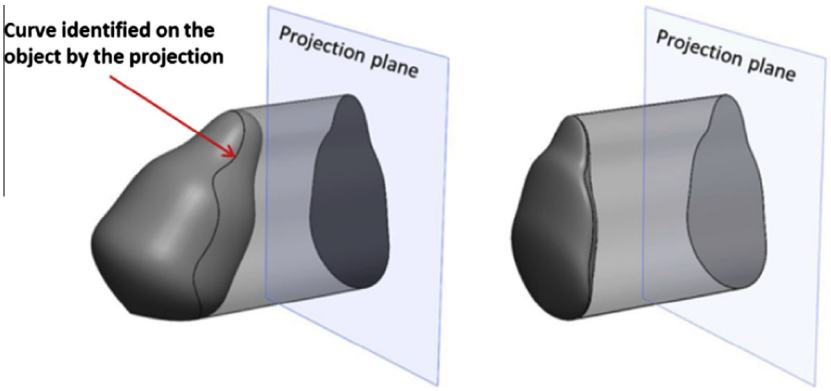

Fig. 5. Curve identified on the object by the projection along the viewing direction.

methodology for 2.5D model retrieval *ab initio* in order to help the reader in getting the complete sense of the method. For this reason a method for obtaining a depth map of the analyzed painting, where subjects are placed in the virtual scene coherently with the perspective, is provided below. A further point to underline is that, differently from similar methods in literature, also oblique planes (i.e. planes represented by trapezoid whose vanishing lines do not converge in the vanishing point) are modeled using the proposed method.

The procedure starts by constructing a Reference Coordinate System (RCS); first, the vanishing point coordinates on the image plane $V = (x_V, y_V)$ are computed [31] thus allowing the definition of the horizon l_h and the vertical line through V, called l_v. Once the vanishing point is evaluated, the RCS is built as follows: the x axis, lying on the image plane and parallel to the horizon (pointing rightwards), the y axis lying on the image plane and perpendicular to the horizon (pointing upwards) and the z axis perpendicular to the image plane (according to the right hand rule). The RCS origin is taken in the bottom left corner of the image plane.

Looking at a generic painted scene with perspective, the following 4 types of planes are then identifiable: *frontal planes*, parallel to the image plane, whose geometry is not modified by the perspective view; *horizontal planes*, perpendicular to the image plane and whose normal is parallel to the y axis (among them it is possible to define the "main plane" corresponding to the ground or floor of the virtual 2.5D scene); *vertical planes*, perpendicular to the image plane and whose normal is parallel to the x axis; *oblique planes*, all the remaining planes, not belonging to the previous three categories. In Fig. 6 some examples of detectable planes from the exemplificative case study are highlighted.

In detail, the identification, and classification, of such planes is performed using the devised GUI with a semi-automatic procedure. First, frontal and oblique planes are selected in image by the user, simply clicking on the appropriate segments of the clustered image. Then, since the V coordinates have been computed, a procedure for automatically classifying a subset of the remaining vertical and horizontal planes starts. In detail, for each cluster C_i, the intersections between the region contained in C_i, and the horizon line are sought. If at least one intersection is found, the plane type is necessarily vertical and, in particular, is vertical left if placed to the left of V and vertical right if placed to the right of V. This statement can be justified by observing that once selected the frontal and/or oblique planes, only horizontal and vertical planes remain and that no horizontal plane can cross the horizon line (they are entirely above or below it). Actually the only horizontal plane (i.e. parallel to the ground) which may "intersect" the horizon line is the one passing through the view point (i.e. at the eye level) and whose representation degenerates in the horizon line itself. If no intersections with the horizon line are found, intersections between the plane and the line passing through the vanishing point and perpendicular to the horizon are sought. If an intersection of this kind is found the plane type is necessarily horizontal (upper horizontal if above V, lower horizontal if below V). Among the detected horizontal planes, it is possible to manually detect the so called "main plane" i.e. a plane taken as a reference for a series of planes intersecting it in the scene. Looking at the most part of renaissance paintings usually such a plane corresponds to the ground (floor). If no intersections with either the horizontal line or with its perpendicular are found, the automatic classification of the plane is not possible; as a consequence, it is requested to manually specify the type of the plane via direct user input.

Once the planes are identified, and classified under one of the above mentioned categories, it is possible to build the virtual flat-layered model by assigning each plane a proper height map. In fact, as widely known, the z coordinate is objectively represented by a gray value: a black value represents the background ($z = 0$), whereas a white value represents the foreground, i.e. the virtual scene element nearest to the observer ($z = 1$).

Since the main plane ideally extends from the foreground to the horizon line, its grayscale representation has been obtained using a gradient represented by a linear graduated ramp extending between two gray levels: the level G_0 corresponding to the nearest point p_0 of the plane in the scene (with reference to an observer) and the level G_1 corresponding to the farthermost point p_1. As a

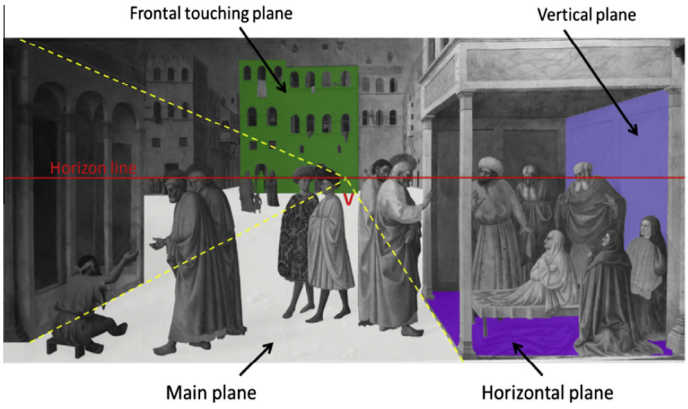

Fig. 6. Some examples of planes detectable from the painted scene.

consequence to the generic point $p \in [p_0, p_1]$ of the main plane is assigned the gray value G given by the following relationship:

$$G = G_0 + (|p - p_0| \cdot S_{grad}) \tag{2}$$

where $S_{grad} = \frac{G_0}{|p_0 - V|}$ is the slope of the linear ramp.

In the 2.5D virtual scene, some planes (which, from now on, are called "touching planes") are recognizable as physically in contact to other ones. These planes share one or more visible contact points: some examples may include a human figure standing on the floor, a picture on a wall, etc. With reference to a couple of touching planes, the first one is called master plane while the second one is called slave plane. The difference between master and slave planes is related to the hierarchical sorting procedure described below. Since contact points between two touching planes should share the same gray value (i.e. same height), considering the main plane and its touching ones (slaves) it is possible to obtain the gray value of a contact point, belonging to a slave plane, directly from the main one (master). This can be done once the main plane has already been assigned a gradient according to Eq. (2). From the inherited gray value $G_{contact}$, the starting gray value G_0 for the slave plane is obtained as follows, in the case of frontal or vertical/horizontal planes respectively:

(a) frontal planes:

$$G_0 = G_{contact} \tag{3}$$

(b) vertical and horizontal planes:

$$\Delta_{grad} = \frac{G_{contact}}{|p_{contact} - V|} \tag{4}$$

$$G_0 = G_{contact} + (|p_{contact} - p_0| \cdot \Delta_{grad}) \tag{5}$$

where $p_{contact}$ is the contact point coordinate ($p \equiv (x, 0)$ for vertical planes, $p \equiv (0, y)$ for horizontal planes). In light of that, the grayscale gradient has to be applied firstly to the main plane so that it is possible to determine the G_0 value for its slave planes. In other words, once plane slaves to the main one have been gradiented, it is possible to find the G_0 values for their own slave planes and so forth. More in detail, assuming the master-slave relationships are available, a hierarchical sorting of the touching planes can be performed. This is possible by observing that the relations between each touching plane can be unambiguously represented by a non-ordered rooted tree, in which the main plane is the tree root node while the other touching planes are the tree leaf nodes. By computing the depth of each leaf (i.e. distance from the root) it is possible to sort the planes according to their depth. The coordinates of the contact point $p_{contact}$ for a pair of touching planes are obtained in different ways depending on their type:

(a) *Frontal planes touching the main plane:* for the generic ith frontal plane touching the main plane, the contact point can be approximated, most of the times, by the lowest pixel h_b of the generic region C_i. Accordingly, the main plane pixel in contact with the considered region, is the one below h_b so that the entire region inherits its gray value. In some special cases, for instance when a scene object is extended below the ground (e.g. a well), the contact point between the two planes may not be the lowest pixel of I_i. In this case it is needed to specify one contact point via user input.

(b) *Vertical planes touching the main plane:* for the generic ith vertical plane (preliminarily classified) touching the main plane it is necessary to determine the region bounding trapezoid. This geometrical construction allows assigning the correct starting gray value (G_0), for the vertical plane gradient, even if the plane has a non-rectangular shape (trapezoidal in perspective view). This is a common situation, also shown in Fig. 7.

By observing Fig. 7b, it is clear that the leftmost pixel of the vertical planar region to be gradiented is not enough in order to identify the point at the same height on the main plane where to inherit G_0 value from. In order to identify such a point, it is actually necessary to compute the vertexes of the bounding box. This step has been carried out using the approach provided in [25].

(c) *Other touching planes:* for every other type of touching plane whose master plane is not the main plane it is necessary to specify both the master plane and one of the contact points shared between the two.

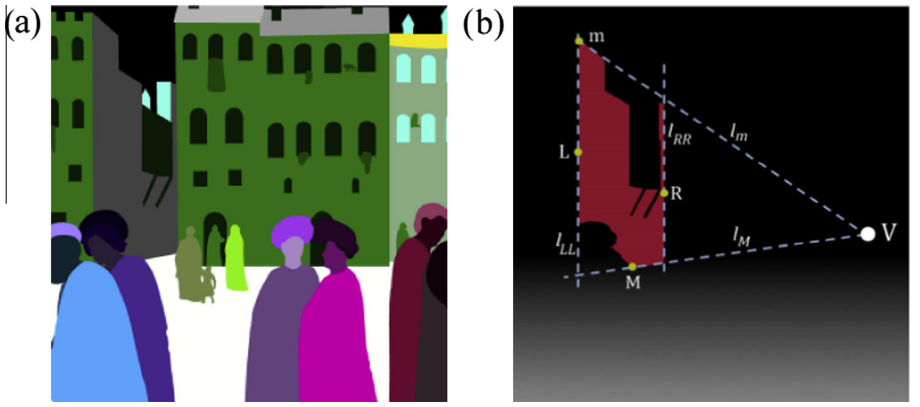

Fig. 7. (a) Detail of the segmented image of "The Healing of the Cripple and the Raising of Tabitha". (b) Example of vertical left plane touching the main plane and its bounding box.

This task is performed via user input since there is no automatic method to univocally determine the contact point.

Using the procedure described above the only planes to be gradiented remains all non-touching planes and oblique ones. In fact, for planes which are not visibly in contact with other planes i.e. non-touching planes (e.g. birds, angels and cherubs suspended above the ground or planes whose contact points are not visible in the scene because hidden by a foreground element), the G_0 value has to be manually assigned by the user choosing, on the main plane, the gray level corresponding to the supposed spatial position of the subject.

Regarding oblique planes, the assignment of the gradient is performed observing that a generic oblique plane can be viewed as an horizontal or vertical plane rotated around one arbitrary axis: when considering the three main axes (x, y, z) three main types of this kind can be identified. Referring to Fig. 8, the first type of oblique planes, called "A", can be expressed as a vertical plane rotated around the x axis while the second (B) is nothing but a horizontal plane rotated around the y axis. The third type of oblique plane (C) can be expressed by either a vertical or horizontal plane, each rotated around the z axis. Of course these three basic rotations can be applied consecutively, generating other types of oblique planes, whose characterization is beyond the scope of this work. Observing the same figure it becomes clear that, except for the "C" type, the vanishing point related to the oblique planes is different from the main vanishing point of the image. For this reason the way to determine the grayscale gradient is almost the same as the other type of planes except for both the V coordinates, that have to be adjusted to an updated value V', and for the gradient direction, that has to be manually specified.

The updated formula for computing the gray level for a generic point P belonging to any type of oblique plane is then:

$$S'_{grad} = \frac{G_0}{|p_0 - V'|} \tag{6}$$

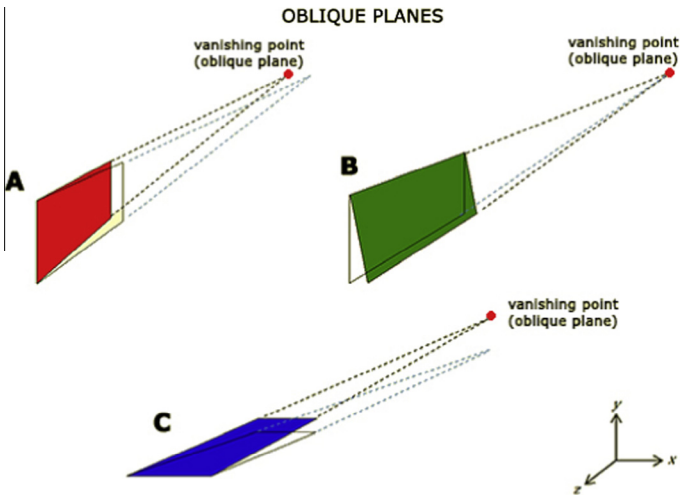

Fig. 8. Characterization of oblique planes.

$$G = G_0 + (|p - p_0| \cdot S'_{grad}) \quad (7)$$

It can be noticed that both x, y coordinates of the points p, p_0 and V' have to be taken into account for calculating the distance, since the sole x or y direction is not representative of the gradient direction when dealing with oblique planes; moreover, both for touching or non-touching oblique planes, the G_0 value has to be manually assigned by the user.

In conclusion, once all the planes are assigned a proper gradient, the final result consists of a grayscale image corresponding to an height map (see Fig. 9). Accordingly, this representation consists of a virtual flat-layered bas-relief.

Of course, this kind of representation performs well for flat or approximately planar regions (see Fig. 10) like, for instance, the wooden panels behind the three figures surrounding the seated woman (Tabitha). Conversely, referring to the Tabitha figure depicted in Fig. 10, the flat-layered representation is only sufficient to represent its position in the virtual scene while its shape (e.g. face, vest etc.) needs to be reconstructed in terms of volume.

2.3. Volumes reconstruction

Once the height map of the scene and the space distribution of the depicted figures have been drafted, it becomes necessary to define the volume of every painted subject, for which is possible, for the viewer, to figure out its actual quasi-three-dimensional shape. As previously stated, in order to accomplish this purpose, it is necessary to translate in shape details all the information elicited from the painting.

First, all objects resembling primitive geometry in the scene are reconstructed using a simple user-guided image processing-based procedure. The user is asked to select the clusters C_i whose final expected geometry is ascribable to a primitive, such as a cylinder or a sphere. Then, since any selected cluster represents a single blob (i.e. a region with constant pixel values) it is easy to detect its geometrical properties like, for instance: centroid, major and minor axis lengths, perimeter and area [32]. On the basis of such values it is straightforward to discriminate between a shape that is approximately circular (i.e. a shape that has to be reconstructed in the form of a sphere) or approximately rectangular (i.e. a shape that has to be reconstructed in the form of a cylinder), using widely known geometric relationships used in blob analysis. After the cluster is classified as a particularly shaped object, a gradient is consistently applied. If the object to be reconstructed is only partially visible in the scene, it is up to the user to manually classify the object. Then, for cylinders user shall select at least two points defining its major axis while, for spheres, he is required to select a point approximately located in the circle center and two points roughly defining the diameter. Once these inputs are provided the grayscale gradient is automatically computed.

Referring to subjects that are not reproducible using primitive geometries (e.g. Tabitha), the reconstruction is mainly performed using SFS-based techniques. This precise choice is due to the fact that the three-dimensional effect of a subject is, usually, realized by the artist using the *chiaroscuro* technique, in order to reproduce on the flat surface of the canvas the different brightness of the real shape under the scene illumination. This leads to consider that, in order to reconstruct the volume of a painted figure, the only useful information that can be worked out from the painting is the brightness of each pixel. Generally speaking, SFS methods prove to be effective for retrieving 3D information (e.g. height map) for synthetic images (i.e. images obtained starting from a given normal map) while the performance of most methods applied to real-world images is still unsatisfactory [33].

Furthermore, as stated in the introductory section, paintings are hand-made artworks and so many aspects of the painted scene (such as silhouette and tones) are unavoidably not accurately reproduced in the image, light direction is unknown in the most of the cases, a diffused light is commonly painted by the artist and imagined surfaces are not perfectly diffusive. These drawbacks make it even more complex the solution of SFS problem with respect to real-world images. For these reasons, the present paper proposes a simplified approach where the final solution i.e. the height map Z_{final} of all the subjects in image is obtained as a combination of three different height maps: (1) "rough shape" Z_{rough}; (2) "main shape" Z_{main} and (3) "fine details shape" Z_{detail}. As a consequence:

$$Z_{final} = \lambda_{rough} Z_{rough} + \lambda_{main} Z_{main} + \lambda_{detail} Z_{detail} \quad (8)$$

It has to be considered that since the final solution is obtained by summing up different contributes, for each of them different simplifying assumptions, valid for

Fig. 9. Final grayscale height-map of "The Healing of the Cripple and the Raising of Tabitha".

Fig. 10. Wooden panels behind the three standing figures are sufficiently modeled using flat representation. Conversely, it is necessary to convey volumetric information of the Tabitha figure.

retrieving each height map, can be stated as explained in the next pages. The main attempt in building a final surface using three different contributes is to reduce the drawbacks due to the non-correct illumination and brightness of the original scene. As explained in detail in the next sections, Z_{rough} is built to overcome possible problems deriving by diffused light in the scene by providing a non-flat surface; Z_{main} is a SFS-based reconstruction obtained using minimization techniques, known to perform well for real-world images (but flattened and over smoothed with respect to the expected surfaces; combining this height map with Z_{rough} the flatten effect is strongly reduced); Z_{detail} is built to reduce the smoothness of both the height maps Z_{rough} and Z_{main}.

2.3.1. Basic principles of SFS

As widely known, under the hypothesis of Lambertian surfaces [22], unique light source set far enough from the scene to assume the light beams being parallel each other and negligible perspective distortion surfaces in a single shaded image can be retrieved by solving the "*Fundamental Equation of SFS*" i.e. a non-linear Partial Derivative Equation (PDE) that express the relation between height gradient and image brightness:

$$\frac{1}{\rho}\Upsilon(x,y)\sqrt{1+|\nabla z|^2}+(l_x,l_y)\nabla z-l_z=0 \tag{9}$$

where $\overrightarrow{L}=[l_x,l_y,l_z]$ is the unit vector opposite to the light direction, ρ is the albedo (i.e. the fraction of light reflected diffusively), $\Upsilon(x,y)$ is the brightness of the pixel located at the coordinates (x,y) and $z(x,y)$ is the height of the retrieved surface.

Among the wide range of methods for solving Eq. (9) [20,26,34,35], minimization methods are acknowledged to provide the best compromise between efficiency and flexibility, since they are able to deliver reasonable results even in case of inaccurate input data, whether caused by imprecise setting (e.g. ambiguous light direction) or inaccuracies of the image brightness. As mentioned above, these issues are unavoidable when dealing with SFS starting from hand-painted images. Minimization SFS approach is based on the hypothesis that the expected surface, which should match the actual one, is (or, at least, is very close to) the minimum of an appropriate functional. Usually the functional, that represents the error to be iteratively minimized between the reconstructed surface and the expected one, is a linear combination of several contributions called "constraints".

Three main different kinds of constraints can be used for solving the SFS problem, according to scientific literature [22]: brightness constraint, which force the final solution to reproduce, pixel by pixel, an image as bright as the initial one; smoothness constraint, which drives the solution towards smooth surface; integrability constraint, which prevents the problem-solving process to provide surfaces that cannot be integrated (i.e. surfaces for which there is an univocal relation between normal map and height).

Brightness constraint is necessary for a correct formulation of the problem, since it is the only one that is based on given data; however it is not sufficient to detect a good solution, due to the "ill-posedness" of the problem: there are indeed infinite scenes that give exactly the input image under the same light condition. For this reason it is necessary to consider, at least, also smoothness and/or integrability constraints so that the minimization process is guided towards a more plausible solution. The downside of adding constraints is that in case of complex surfaces characterized by abrupt changes of slope and high frequency details the error in solution caused by smoothness or integrability constraints becomes relevant. This particular effect, called over-smoothing error, leads the resolution algorithm to produce smoother surfaces respect to the expected one leading to a loss in fine surface details.

Another important aspect to be taken into account, especially if dealing with paintings representing open-air scenarios, is that the possible main light source is sometimes combined with a fraction of diffused reflected light, usually called ambient light. Consequently, the image

results more artistically effective, but unfortunately also affected by lower contrast resolution with respect to a single-spot illumination based scene. The main problem arising from the presence of diffused light is that in case volumetric shape is retrieved using SFS without considering it, the result will appear too "flat" due to the smaller brightness range available in image (deepness is loosed in case of low frequency changes in slope). The higher is the contribute of the diffused light with respect to spot lights, the lower is the brightness range available for reconstruction and so the flatter is the retrieved volume.

In order to reduce this effect, a method could be to subtract from each pixel of the original image Υ a constant value L_d, so that the tone histogram of the new obtained image Υ' is left-shifted. As a consequence, fundamental equation of SFS is to be changed using Υ' instead of Υ i.e.:

$$\rho(\vec{N}\cdot\vec{L})+L_d=\Upsilon' \tag{10}$$

Eq. (10) could be solved using minimization techniques too. Unfortunately, the equation is mathematically verified only when dealing with real or synthetic scenes. For paintings considering the fraction of diffused light as a constant is a mere approximation and performing a correction of the original image Υ with such a constant value cause an unavoidable loss in image details. For this reason a method for taking into account diffused ambient light thus allowing to avoid over-flattening effect in shape reconstruction is required. This is the reason why in the present work a rough shape surface is retrieved using the approach provided in the following section.

2.3.2. Rough shape Z_{rough}

As mentioned above, rarely the scene painted by an artist can be treated as a single illuminated scene. As a consequence, in order to come up with the over-flattening effect due to the possible presence of ambient light, L_d, in the present work the gross volume of the final shape is obtained by using inflating and smoothing techniques, applied to the silhouette of every figure represented in the image.

The proposed procedure is composed of two consecutive operations: rough inflating and successive fine smoothing. Rough inflating is a one-step operation that provides a first-tentative shape, whose height map Z_{inf} is, pixel by pixel, proportional to the Euclidean distance from the outline. This procedure is quite similar to the one used in commercial software packages like ArtCam. However, as demonstrated in [26] the combination of rough inflating and successive smoothing produces better results.

Moreover, the operation results extremely fast but, unfortunately, the obtained height map Z_{inf} is not acceptable as it stands, appearing very irregular and indented: this inconvenience is caused by the fact that the digital image is discretized in pixels, producing an outline that appears as a broken line instead of a continuous curve. For this reason the height map Z_{inf} is smoothed by iteratively applying a short radius (e.g. sized 3×3) average filter to the surface height map keeping the height value on the outline unchanged.

The result of this iterative procedure is an height map Z_{rough} defining, in its turn, a surface resembling the rough shape of the subject; the importance of this surface is twofold: first it is one of the three contributes that linearly combined will provide the final surface; secondly it is used as initialization surface for obtaining the main shape Z_{main} as explained in the next section.

In Fig. 11 the surface obtained for Tabitha figure is depicted, showing visual effect similar to the one obtained by inflating a balloon roughly shaped as the final object to be reconstructed.

It is important to highlight that Z_{rough} is built independently from the scene illumination, still allowing to reduce the over-flattening effect. The use of this particular method allows to avoid the use of Eq. (10) thus simplifying the SFS problem i.e. to neglect diffuse illumination L_d. Obviously, the obtained surface results excessively smoothed, with respect to the expected one. This is the reason why, as explained later, the fine detail surface is retrieved.

2.3.3. Main shape Z_{main}

In this phase a surface, called "main surface" is basically retrieved using the SFS-based method described in [7], here quickly reminded to help the reader in understanding the overall procedure. Differently from the cited method, however, in this work a method for determining unknown light source $\vec{L}=[l_x,l_y,l_z]$ is implemented. This is a very important task for reconstructing images from paintings since the principal illumination of the scene can be only be guessed by the observer (and in any case empirically measured, since it is not a real or a synthetic scene).

First a functional to be minimized is built as a linear combination of brightness (B) and smoothing (S) constraints, into the surface reconstruction domain D:

$$E=B+\lambda S=\sum_{i\in D}\left(1/\rho G_i-\vec{N}_i^T\cdot\vec{L}\right)^2+\lambda\sum_{\{ij\}\in D}\left(\vec{N}_i-\vec{N}_j\right)^2 \tag{11}$$

where i is the pixel index; j is the index of a generic pixel belonging to the 4-neighborhood of pixel i; G_i is the brightness of pixel i (range [0–1]); $\vec{N}_i=[n_{i,x},n_{i,y},n_{i,z}]$, $\vec{N}_j=[n_{j,x},n_{j,y},n_{j,z}]$ are the unit length vectors normal to the surface (unknown) in positions i and j, respectively; λ is a regularization factor for smoothness constraint (weight).

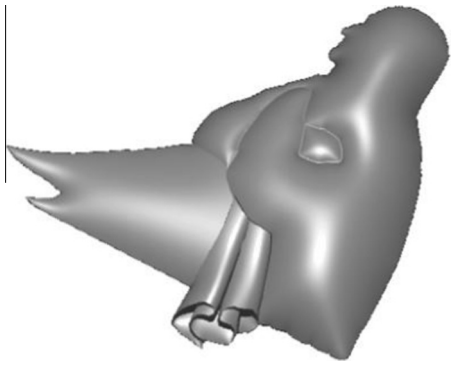

Fig. 11. Height map Z_{rough} obtained using inflating and successive iterative smoothing on the Tabitha figure.

Thanks to the proper choice of the minimization unknown, the outcome results to be a quadratic form. As a consequence its gradient is linear and the minimization process can be carried out indirectly by optimizing the functional gradient itself (Eq. (9)), by using Gauss-Seidel with Successive Over Relaxation (SOR) iterative method, which is proved to allow a very fast convergence to the optimized solution [20,24].

$$\begin{cases} \min_{\Phi} E = \min_{\Phi} \frac{1}{2}\Phi^T A\Phi + \Phi^T b + c \rightarrow \nabla(E) = A\Phi + b = 0 \\ \Phi = [n_{1,x}, \ldots, n_{k,x}, n_{1,y}, \ldots, n_{k,y}, n_{1,z}, \ldots, n_{k,z}]^T \end{cases} \tag{12}$$

where k is the overall number of pixel defining the shape to be reconstructed.

However, Eq. (12) can be built only once the vector $\overrightarrow{L} = [l_x, l_y, l_z]$ is evaluated. For this reason the devised GUI implements a simple user-driven tool whose aim is to quickly and easily detect the light direction in the image. Despite several approaches exist to cope with light direction determination on real-world or synthetic scenes [36], these are not adequate for discriminating scene illumination on painted subjects. Accordingly, a user-guided procedure has been setup.

In particular, the user can regulate light unit-vector components along axes x, y and z, by moving the corresponding slider on a dedicated GUI (see Fig. 12), while the shading generated by using such illumination is displayed on a spherical surface. If a shape on the painting is locally approximated by a sphere, then the shading obtained on the GUI needs to be as close as possible to the one on the picture. This procedure makes it easier for the users to properly set the light direction. Obviously this task requires user to guess the scene illumination on the basis of an analogy with the illumination distribution on a known geometry.

Once the matrix form of the functional has been correctly built, it is possible to set different kinds of boundary conditions. This step is the crucial point of the whole SFS procedure, since it is the only operation that allows the user to properly guide the automatic evaluation of the geometry itself. Two main kinds of boundary conditions have to be set: the first drives the reconstruction by fixing the unknowns on the silhouette outline (*Silhouette Boundary Condition, SBC*) while the second by fixing these unknowns on the selected white areas outlines (*Morphological Boundary Condition, MBC*), both performed by setting interactively as the local maxima or minima height points of the expected surface [37].

More in depth, when the subject to be reconstructed is clearly detached from the background, the *SBC* allows to discriminate between concave and convex global shape. This is obtained by imposing the unit-normal inward or outward pointing respect to the silhouette itself. In addition, since background unknowns are not included in the surface reconstruction domain D, they are automatically forced to lie on the z-axis as explained in Eq. (13):

$$\overrightarrow{N}^T_{b \notin D} = [0, 0, 1]^T \tag{13}$$

MBC, instead, plays a fundamental role to locally overcome the concave–convex ambiguity. In particular, users are required to specify, for a number of white regions in the original image (possibly for all of them), which ones correspond to local maxima or minima (in terms of surface height) figuring out the final shape as seen from an observer located in correspondence of the light source. Once such points are selected, the algorithm provided in [99] is used to evaluate the normals to be imposed.

In addition, unknowns $\overrightarrow{N}_w$ coinciding with white pixels w included in D are automatically set equal to $\overrightarrow{L}$:

$$\forall w \in D \mid I_w = 1 \rightarrow \overrightarrow{N}_w = \overrightarrow{L} \tag{14}$$

At the end of this procedure, the matrix formulation of the problem results modified and reduced, properly guiding the successive minimization phase:

$$\nabla(E_r) = A_r\hat{\Phi} + b_r \tag{15}$$

where E_r, A_r and b_r are, respectively, a reduced version of E, A and b.

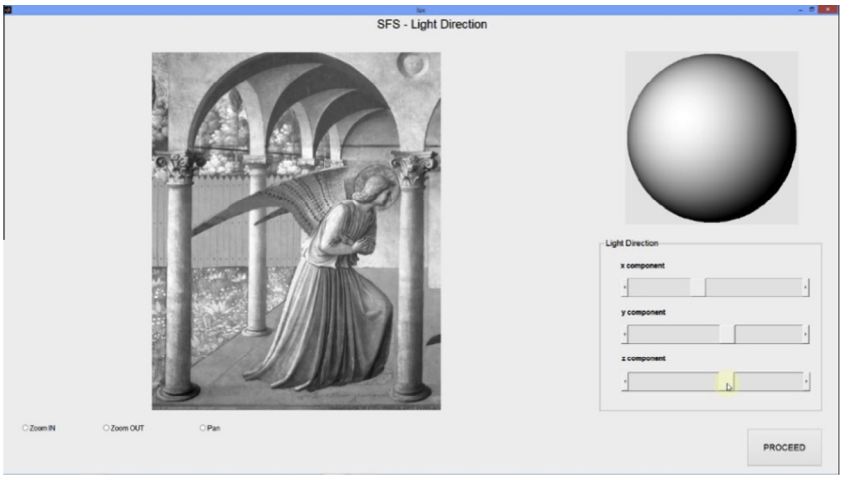

Fig. 12. GUI implemented to set the light unit-vector components along axes x, y and z.

The minimization procedure can provide a more reliable solution if the iterative process is guided by initial guess of the final surface. In such terms, the surface obtained from Z_{rough} is effectively used as initialization surface. Once the optimal normal map has been evaluated, the height map Z_{main} is obtained by minimizing the difference between the relative height, $z = z_i - z_j$, between adjacent pixels and a specific value, q_{ij}, that express the same relative height calculated by fitting an osculating arc between the two unit-normals:

$$E_2 = \sum_{\{i,j\}} \left((z_i - z_j) - q_{ij} \right)^2 \tag{16}$$

The final result of this procedure (see Fig. 13) consists of a surface Z_{main} roughly corresponding to the original image without taking into account diffuse light. As already stated above, such a surface results excessively flat, with respect to the expected one. This is, however, an expected solution since, as reaffirmed above, the over-flattening is corrected by mixing together Z_{main} and Z_{rough}.

In other words, combining Z_{main} with Z_{rough} is coarsely equivalent to solve SFS problem using both diffuse illumination and principal illumination thus allowing the reconstruction of a surface effectively resembling the original (imagined) shape of the reconstructed subject and avoiding over-flattening. The advantage here is that any evaluation of the diffused light is required since Z_{rough} is built independently from light in image.

Obviously, the value of λ_{rough} and λ_{main} have to be properly set in order to balance the effect of inflating (and smoothing) with the effect of SFS-based reconstruction. Moreover finest details in the scene are not reproduced since both techniques tend to smooth down the surface (smoothing procedure in retrieving Z_{rough} plus over-smoothing due to the presence of smoothing constraint in Z_{main} SFS-based reconstruction). This is the reason why another height map (and derived surface) has to be computed as explained in the next section.

2.3.4. Fine details shape Z_{detail}

Inspired by some known techniques that commonly available in both commercial software packages and in literature works [38], a simple, but efficient, way to reproduce the fine details is to consider the brightness of the input image as a height map (see Fig. 14). This means that the Z_{detail} height map is provided by the following equation:

$$Z_{detail}(i,j) = \frac{1}{\rho_k} \Upsilon(i,j) \tag{17}$$

where:

- $Z_{detail}(i,j)$ is the height map value at the pixel of coordinate (i,j).
- $Y(i,j)$ is the brightness value of the pixel (i,j) in the image Y.
- ρ_k is the value of the albedo of the segment k of the image, to which the pixel belongs.

The obtained height map is not the actual one since object details perpendicular to the scene principal (painted) light will result, in this reconstruction, closer to the observer even if they are not; moreover convexity and concavity of the subjects are not discerned. However the obtained shape visually resembles the desired one and thus can be used to improve the overall surface retrieval.

As an alternative, the height map Z_{detail} can be obtained by the magnitude of the image gradient according to the following equation which, for a given pixel with coordinates (i,j), can be approximately computed, for instance, with reference to its 3×3 neighborhood according to equation 19 making use of image gradient.

$$Z_{detail}(i,j) = \left| \nabla \left(\frac{1}{\rho_k} \Upsilon \right) \right|_{(i,j)} \tag{18}$$

$$\left| \nabla \left(\frac{1}{\rho_k} \Upsilon \right) \right| = \sqrt{\nabla_x^2 + \nabla_y^2} \tag{19}$$

where:

$$\nabla_x \cong \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \otimes \Upsilon \text{ and } \nabla_y \cong \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \otimes \Upsilon \tag{20}$$

Fig. 13. Height map Z_{main} obtained for Tabitha figure by solving the SFS problem with minimization algorithm and using Z_{rough} as initialization surface.

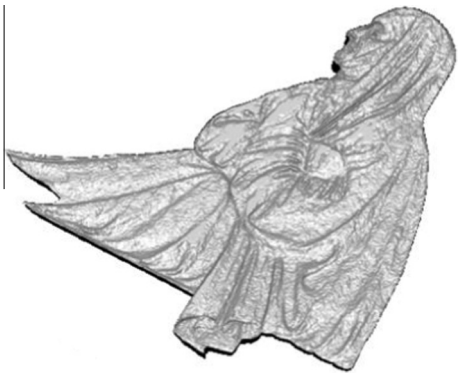

Fig. 14. Height map Z_{detail} obtained for Tabitha figure using the original grayscale image as a height map.

Fig. 15. Height map Z_{detail} obtained for Tabitha figure using the gradient of the original image as a height map.

It should be noticed that not even the height map obtained by Eq. (18) is correct: brighter (though more in relief) regions are the ones where a sharp variation in the original image occurs, thus not corresponding to the regions actually closer to the observer (Fig. 15). However, also in this case the retrieved surface visually resembles the desired one; for this reason it can be used to improve the surface reconstruction.

Accordingly, the height map obtained using one of the two formulations provided in Eqs. (17) or (19) is worthwhile only setting the weight in Eq. (8) to a value lower than 0.05. In fact, the overall result obtainable by combining Z_{detail} with Z_{main} and Z_{rough} using such a weight is much more realistically resembling the desired surface.

2.4. Virtual bas-relief reconstruction

Using the devised GUI it is possible to interactively combine the contributions of *perspective-based reconstruction*, *SFS*, *rough* and *fine detail surfaces*. In particular it is possible to decide the optimum weights for the different components and to assess how much relief is given to each object with respect to the others. Users are allowed to select the weight by means of relative sliders, showing in real-time the aspect of the final surface (see Fig. 16). In this phase, the system prevents the user from generating, for a given object, a surface whose relief exceeds the one of other objects closer to the observer.

In Fig. 17 the virtual 2.5D model (bas-relief) obtained by using the devised procedure is depicted. In order to appreciate the differences between the perspective geometry-based reconstruction and the volumes-based one, in Fig. 17 the model is split into two parts. On the left part the flattened bas-relief obtained from the original image is shown to illustrate the position of subjects in the scene, but without volumes. On the right part, the complete reconstruction is provided (both position in the virtual scene and volumetric information).

2.5. Rapid prototyping of the virtual bas-relief

Once the final surface has a satisfying quality, the procedure allows the user to produce a STL file, ready to be manufactured with reverse prototyping techniques or CNC milling machines.

In Fig. 18, the CNC prototyped model for "The Healing of the Cripple and the Raising of Tabitha" is shown. The physical prototype is sized about 900 mm × 400 mm × 80 mm. In the same figure, a detail from the final prototype is also showed.

3. Case studies

The devised method was widely shared with both the Italian Union of Blind and Visually Impaired People in Florence (Italy) and experts working in the Cultural Heritage field, with particular mention to experts from the Musei Civici Fiorentini (Florence Civic Museums, Italy) and from Villa la Quiete (Florence, Italy). According to their suggestions, authors realized a wide range of bas-reliefs of well-known artworks of the Italian Renaissance including "The Annunciation" of Beato Angelico (see Fig. 19) permanently

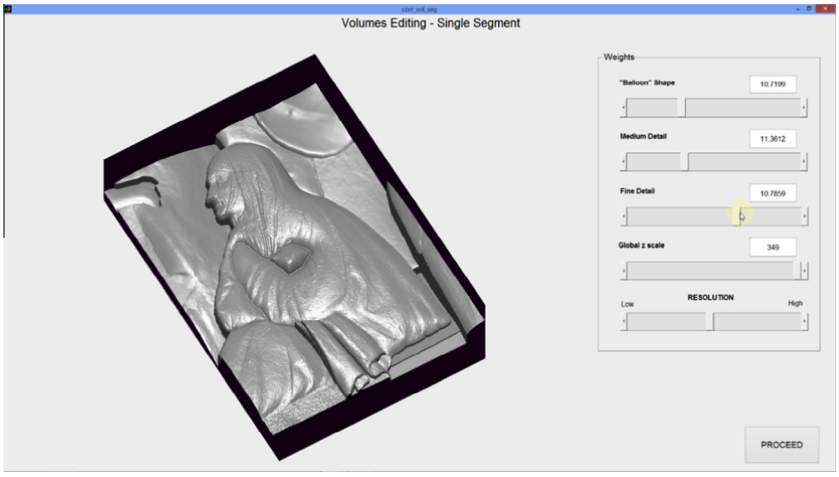

Fig. 16. GUI devised for interactively combining the contributions from perspective-based reconstruction, SFS, Inflated and Fine detail surfaces.

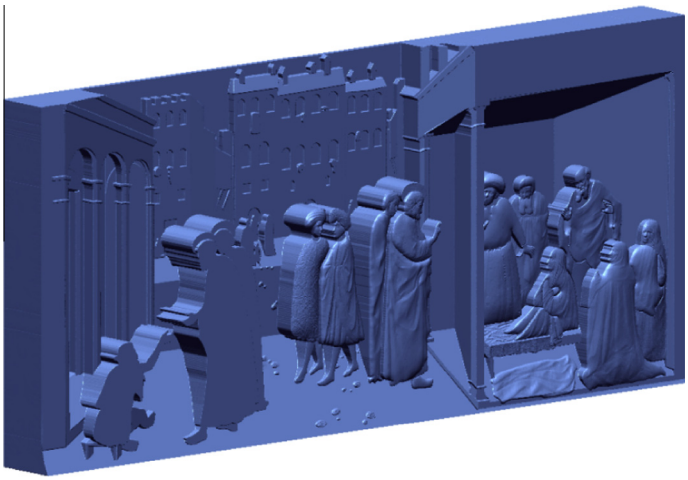

Fig. 17. 2.5D virtual model obtained by applying the proposed methodology on the image of Fig. 2. On the left part of the model, the flattened bas-relief obtained from the original image is shown; on the right part of the model the complete reconstruction is provided.

Fig. 18. Prototype of "The Healing of the Cripple and the Raising of Tabitha" obtained by using a CNC.

displayed at the Museo di San Marco (Firenze, Italy), some figures from the "Mystical marriage of Saint Catherine" by Ridolfo del Ghirlandaio (see Fig. 20) and the "Madonna with Child and angels" by Niccolò Gerini (see Fig. 21), both displayed at Villa La Quiete (Firenze, Italy).

4. Discussion

In cooperation with the Italian Union of Blind and Visually Impaired People in Firenze (Italy), a panel of 14 users with a total visual deficit (8 congenital (from now on CB) and 6 acquired (from now on AB)), split in two age groups (between 25–47 years -7 users- and between 54–73 -7 users) has been selected for testing the realized models, using the same approach described in [4]. The testing phase was performed by a professional who is specifically trained to guide people with visual impairments in tactile exploration.

For this purpose, after a brief description about the general cultural context, the pictorial language of each artist and the basic concept of perspective view (i.e. a description of how human vision perceive objects in space) the interviewees were asked to imagine the position of subjects in the 3D space and their shape on the basis of a two-levels tactile exploration: completely autonomous and guided. In the first case, people from the panel were asked to provide a description of the perceived subjects, their position in space (both absolute and mutual) and to guess their

Fig. 19. Prototype of "The Annunciation" of Beato Angelico permanently placed in the upper floor of the San Marco Museum (Firenze), next to the original Fresco.

Fig. 20. Prototype resembling the Maddalena and the Child figures retrieved from the "Mystical marriage of Saint Catherine" by Ridolfo del Ghirlandaio.

shape after a tactile exploration without any interference from the interviewer and without any limitation in terms of exploration time and modalities. In the second phase, the expert provided a description of the painted scene including subjects and mutual position in the (hypothetic) 3D space. After this explanation, the panel was asked to identify again described subjects and their position in the virtual 2.5D space.

Considering the four tactile bas-relief models, about 50% of AB and 37.5% of CB was able to perceive the main subjects of the scene in the correct mutual position during completely autonomous exploration. As a consequence about 43% of the panel was able to properly understand the reconstructed scene. Moreover, after the verbal description provided in the second exploration phase, 86% of the panel proved to be capable of providing a sufficiently clear description of the touched scene.

Furthermore, after the second exploration phase, a closed-ended question ("why the model is considered readable") was administered to the panel; the available answers were: (1) "better perception of depth", (2) "better perception of shapes", (3) "better perception of the subjects mutual position", (4) "better perception of details". Results, depicted in Fig. 22, depend on the typology of visual disability. AB people believed that the reconstructed model allows, primarily, a better perception of position (50%) and, secondarily, a better perception of shapes (33%). Quite the reverse, as many as the 50% of the CB people declared to better perceive the shapes.

Despite the complexity in identifying painted subjects, and their position, from a tactile bas-relief, and knowing that a single panel of 14 people is not sufficient to provide a reliable statistics on this matter, it can be qualitatively stated that perspective-based reconstruction 2.5D models from painted images are, in any case, an interesting attempt in enhancing the artworks experience of blind and visually impaired people. Accordingly a deeper work in this field is highly recommended.

Fig. 21. Prototype realized for "Madonna with Child and angels" by Niccolò Gerini.

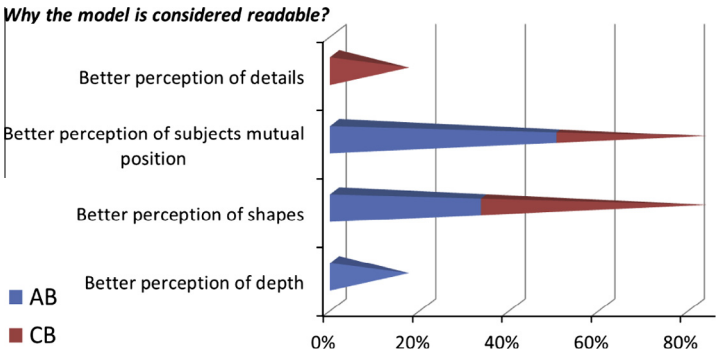

Fig. 22. Percentages of AB (blue) and CB people (red) related to the reasons of their preference. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Typical issues of SFS applied to hand drawings and proposed solutions.

Cause	Effect using SFS methods	Solution	Possible drawbacks
Presence in the scene of a painted (i.e. guessed by the artist) diffuse illumination	Flattened final surface	Retrieval of a rough surface Z_{rough} obtained as a result of inflating and smoothing iterative procedure. This allows to solve the SFS problem using only the principal illumination	Loss of details in the reconstructed surfaces. Too much emphasis on coarse volumes
Incorrect shading of objects due to artistic reproduction of subjects in the scene	Errors in shape reconstruction	Retrieval of a surface Z_{main} based on SFS modified method using minimization and a set of boundary conditions robust to the possible errors in shading	If the shading of painted objects is grossly represented by the artist, the reconstruction may appear unfaithful
Incorrect scene principal illumination	Combined with the incorrect shading, this leads to errors in shape illumination	Use of an empirical method for determining, approximately, the illumination vector. Combining Z_{main} with Z_{rough} is equivalent to solve SFS problem using both diffuse illumination and principal illumination	Solving the method with more than one principal illumination leads to incorrect reconstructions
Loss of details due to excessive inflating and to over smoothing effect in SFS-based reconstruction (e.g. using high values of smoothness constraint)	None	Use of a refinement procedure allowing to take into account finer details of the reconstructed objects	None

5. Conclusions

The present work presented an user-guided orderly procedure meant to provide 2.5D tactile models starting from single images. The provided method deals with a number of complex issues typically arising from the analysis of painted scenes: imperfect brightness rendered by the artist to the scene, incorrect shading of subjects, incorrect diffuse light, inconsistent perspective geometry. In full awareness of these shortcomings, the proposed method was intended to provide a robust reconstruction by using a series of reasoned assumptions while contemporary being tolerant to non-perfect reconstruction results. As depicted in Table 1, most of the problems arising from 2.5D reconstruction starting from painted images are confronted with and, at least partially, solved using the proposed method. In these terms, the method is intended to add more knowledge, and tools, dedicated to the simplification of 2D to 2.5D translation of artworks, making more masterpieces available for visually impaired people and allowing a decrease of final costs for tactile paintings creation.

With the aim of encouraging further works in this field a number of open issues and limitations can be drafted. Firstly, since the starting point of the method consists of image clustering, the development of an automatic and accurate segmentation of painting subjects could dramatically improve the proposed method; extensive testing to assess the performance of state of the art algorithms applied to paintings could be also helpful for speeding up this phase. Secondly, the reconstruction of morphable models of parts commonly represented in paintings (such as hands, limbs or also entire human bodies) could be used in order to facilitate the relief generation. Using such models could avoid solving complex SFS-based algorithms for at least a subset of subjects. Some improvements could also be performed (1) by developing algorithms aimed to automatically suppress/compress useless parts (for instance the part of the scene comprised between the background and the farthest relevant figure which needs to be modeled in detail) and (2) to perform automatic transitions between adjacent areas (segments) with appropriate constraints (e.g. tangency). The generation of slightly undercut models, to facilitate the comprehension/recognition of the

represented objects and to make the exploration more enjoyable, is another open issue to be well-thought-out in the next future. Finally, future works could be addressed to embrace possible inputs and suggestions from neuroscientists in order to make the 2.5D reproductions more "effective" so that it is possible to improve the "aesthetic experience" of the end-users. Moreover, the possibility of enriching the tactile exploration experience by developing systems capable of tracing users fingers and of providing real time audio feedback could be investigated in the next future.

Acknowledgments

The authors wish to acknowledge the valuable contribution of Prof. Antonio Quatraro, President of the Italian Union of Blind People (UIC) Florence (Italy), in helping the authors in the selection of artworks and in assessing the final outcome. The authors also wish to thank the Tuscany Region (Italy) for co-funding the T-VedO project (PAR-FAS 2007-2013), which originated and made possible this research, the Fondo Edifici di Culto – FEC of the Italian Ministry of Interior (Florence, Italy) and the Carmelite Community.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.gmod.2014.10.001.

References

[1] R.L. Klatzky, S.J. Lederman, C.L. Reed, There's more to touch than meets the eye: the salience of object attributes for haptics with and without vision, J. Exp. Psychol. Gen. 116 (1987) 356–369.
[2] A. Streri, E.S. Spelke, Haptic perception of objects in infancy, Cogn. Psychol. 20 (1) (1988) 1–23.
[3] A. Reichinger, M. Neumüller, F. Rist, S. Maierhofer, W. Purgathofer, Computer-aided design of tactile models – taxonomy and case studies, in: K. Miesenberger, A. Karshmer, P. Penaz, W. Zagler (Eds.), Computers Helping People with Special Needs. 7383 Volume of Lecture Notes in Computer Science, Springer, Berlin/Heidelberg, 2013, pp. 497–504.
[4] M. Carfagni, R. Furferi, L. Governi, Y. Volpe, G. Tennirelli, Tactile representation of paintings: an early assessment of possible computer based strategies, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7616 LNCS, 2012. pp. 261–270.
[5] L. Thompson, E. Chronicle, Beyond visual conventions: rethinking the design of tactile diagrams, Br. J. Visual Impairment 24 (2) (2006) 76–82.
[6] S. Oouchi, K. Yamazawa, L. Secchi, Reproduction of tactile paintings for visual impairments utilized three-dimensional modeling system and the effect of difference in the painting size on tactile perception, Computers Helping People with Special Needs, Springer, 2010. pp. 527–533.
[7] Y. Volpe, R. Furferi, L. Governi, G. Tennirelli, Computer-based methodologies for semi-automatic 3D model generation from paintings, Int. J. Comput. Aided Eng. Technol. 6 (1) (2014) 88–112.
[8] Y. Horry, K. Anjyo, K. Arai, Tour into the picture: using a spidery mesh interface to make animation from a single image, in: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, ACM Press/Addison-Wesley, 1997.
[9] D. Hoiem, A.A. Efros, H. Martial, Automatic photo pop-up, ACM Transactions on Graphics (TOG), vol. 243, ACM, 2005.
[10] J. Wu, R.R. Martin, P.L. Rosin, X.-F. Sun, Y.-K. Lai, Y.-H. Liu, C. Wallraven, of non-photorealistic rendering and photometric stereo in making bas-reliefs from photographs, Graphical Models 76 (4) (2014) 202–213. ISSN 1524-0703, http://dx.doi.org/10.1016/j.gmod.2014.02.002.
[11] J. Wu, R. Martin, P. Rosin, X.-F. Sun, F. Langbein, Y.-K. Lai, A. Marshall, Y.-H. Liu, Making bas-reliefs from photographs of human faces, Comput. Aided Des. 45 (3) (2013) 671–682.
[12] Z.K. Huang, X.W. Zhang, W.Z. Zhang, L.Y. Hou, A new embossing method for gray images using Kalman filter, Appl. Mech. Mater. 39 (2011) 488–491.
[13] W. Song, A. Belyaev, H.-P. Seidel, Automatic generation of bas-reliefs from 3d shapes, shape modeling and applications, 2007. SMI'07. In: IEEE International Conference on. IEEE, 2007.
[14] Sourin, Alexei. Functionally based virtual computer art, in: Proceedings of the 2001 Symposium on Interactive 3D Graphic, ACM, 2001.
[15] Z. Li, S. Wang, J. Yu, K.-L. Ma, Restoration of brick and stone relief from single rubbing images, Visualization Comput. Graphics IEEE Trans. 18 (2012) 177–187.
[16] M. Wang, J. Chang, J.J. Zhang, A review of digital relief generation techniques, Computer Engineering and Technology (ICCET), in: 2010 2nd International Conference on, IEEE, 2010, p. 198–202.
[17] T. Weyrich et al., Digital bas-relief from 3D scenes. ACM Transactions on Graphics TOG, vol. 26(3), ACM, 2007.
[18] J. Kerber, Digital Art of Bas-relief Sculpting, Master's Thesis, Univ. of Saarland, Saarbrücken, Germany, 2007.
[19] A. Reichinger, S. Maierhofer, W. Purgathofer, High-quality tactile paintings, J. Comput. Cult. Heritage 4 (2) (2011). Art. No. 5.
[20] P. Daniel, J.-D. Durou, From deterministic to stochastic methods for shape from shading, in: Proc. 4th Asian Conf. on Comp. Vis, Citeseer, 2000, pp. 1–23.
[21] L. Di Angelo, P. Di Stefano, Bilateral symmetry estimation of human face, Int. J. Interactive Design Manuf. (IJIDeM) (2012) 1–9. http://dx.doi.org/10.1007/s12008-012-0174-8.
[22] J.-D. Durou, M. Falcone, M. Sagona, Numerical methods for shape-from-shading: a new survey with benchmarks, Comput. Vis. Image Underst. 109 (2008) 22–43.
[23] R.T. Frankot, R. Chellappa, A method for enforcing integrability in shape from shading algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 10 (4) (1988) 439–451.
[24] T.-P. Wu, J. Sun, C.-K. Tang, H.-Y. Shum, Interactive normal reconstruction from a single image, ACM Trans. Graphics (TOG) 27 (2008) 119.
[25] R. Furferi, L. Governi, N. Vanni, Y. Volpe, Tactile 3D bas-relief from single-point perspective paintings: a computer based method, J. Inform. Comput. Sci. 11 (16) (2014) 1–14. ISSN: 1548-7741.
[26] L. Governi, M. Carfagni, R. Furferi, L. Puggelli, Y. Volpe, Digital bas-relief design: a novel shape from shading-based method, Comput. Aided Des. Appl. 11 (2) (2014) 153–164.
[27] L.G. Brown, A survey of image registration techniques, ACM Comput. Surveys (CSUR) 24 (4) (1992) 325–376.
[28] M.W. Schwarz, W.B. Cowan, J.C. Beatty, An experimental comparison of RGB, YIQ, LAB, HSV, and opponent color models, ACM Trans. Graphics (TOG) 6 (2) (1987) 123–158.
[29] E. Nadernejad, S. Sharifzadeh, H. Hassanpour, Edge detection techniques: evaluations and comparisons, Appl. Math. Sci. 2 (31) (2008) 1507–1520.
[30] W.A. Barrett, E.N. Mortensen, Interactive live-wire boundary extraction, Med. Image Anal. 1 (4) (1997) 331–341.
[31] C. Brauer-Burchardt, K. Voss, Robust vanishing point determination in noisy images, Pattern Recognition, in: Proceedings 15th International Conference on, vol. 1, IEEE, 2000.
[32] J.R. Parker, Algorithms for Image Processing and Computer Vision, John Wiley & Sons, 2010.
[33] O. Vogel, L. Valgaerts, M. Breuß, J. Weickert, Making shape from shading work for real-world images, Pattern Recognition, Springer, Berlin Heidelberg, 2009. pp. 191–200.
[34] R. Huang, W.A.P. Smith, Structure-preserving regularisation constraints for shape-from-shading, Comput. Anal. Images Patterns Lecture Notes Comput. Sci. 5702 (2009) 865.
[35] P.L. Worthington, E.R. Hancock, Needle map recovery using robust regularizers, Image Vis. Comput. 17 (8) (1999) 545–557.
[36] C. Wu et al., High-quality shape from multi-view stereo and shading under general illumination, Computer Vision and Pattern Recognition (CVPR), in: 2011 IEEE Conference on. IEEE, 2011.
[37] L. Governi, R. Furferi, L. Puggelli, Y. Volpe, Improving surface reconstruction in shape from shading using easy-to-set boundary conditions, Int. J. Comput. Vision Robot. 3 (3) (2013) 225–247.
[38] K. Salisbury et al., Haptic rendering: Programming touch interaction with virtual objects, in: Proceedings of the 1995 symposium on Interactive 3D graphics, ACM, 1995.