TOWARDS A 3D MULTISPECTRAL SCANNER: AN APPLICATION TO MULTIMEDIA

Alamin Mansouri, Alexandra Lathuilière, Franck S. Marzani, Yvon Voisin, and Pierre Gouton

1 Le2i, UMR CNRS 5158, Université de Bourgogne, Aile des sciences de l’ingénieur, BP 47870 Dijon CEDEX, France
{alamin.mansouri, franck.marzani, pgouton}@u-bourgogne.fr

2 Le2i, UMR CNRS 5158, Université de Bourgogne, IUT du Creusot, 12 rue de la fonderie, 71200 Le Creusot, France
{a.lathuiliere, y.voisin}@iutlecreusot.u-bourgogne.fr

Abstract. We describe a new stereoscopic system based on a multispectral camera and an LCD-Projector. The novel concept we want to show consists in the use of multispectral information for 3D-objects reconstruction. Each 3D-point is linked to a curve representing the spectral reflectance. This latter is a physical representation of the matter and presents the advantage over color information in that sense it is independent from illuminant, observer, and acquisition device. This knowledge allows some multimedia applications like simulating illumination change. We exploit the proposed system in the case of 3D artwork objects. We first present an easy methodology to geometrically and spectrally calibrate such a system. We then describe an algorithm for recovering 3D coordinates based on triangulation and an algorithm for reflectance curves reconstruction based on neural networks. The results are encouraging; they confirm the feasibility of such a system.

Keywords: Multispectral imaging, 3D Reconstruction, Neural networks, Illuminant change simulation, Multimedia applications, Virtual museum.

1 Introduction

In the field of 3D object reconstruction and metrology, vision-based systems are becoming increasingly prevalent, including industrial applications. In particular, within the framework of applications whose object has a significant volume, they seem preferable to techniques based on interferometry or the watered effect, techniques which mainly measure nanometer order depths.

Vision systems can be divided into two categories: passive or active vision. The systems included in the first category use only cameras to acquire the 3D object. One can use one or several cameras to acquire the same object, but it is always necessary to have at least two images of the same object acquired from different angle of view. The method used to give position and depth information is based on the matching of the pixels between images in order to reconstruct by triangulation their 3D position. The problem with such methods is in the detection and matching of the characteristic points. It is a non trivial task: low textured objects become very difficult to analyze because of the lack of characteristic points appearing on their surface. In the vision systems called Active, one passive sensor could be replaced by a device which projects a structured light onto the object. This light can be a laser beam, but it is necessary to sweep the object, or it can be an LCD-projector, in this case one projection can be sufficient. The structured light creates a kind of texture on the object surface that a camera can then acquire. If it is supposed that this system is geometrically calibrated, the position and the depth of the object points, as illuminated by the projected pattern, can be calculated. Many types of structured light have already been studied. The interested
reader can refer to the Battle article [1] in which one can find a state of the art techniques.

Within this framework, the use of systems made of a camera and an LCD-projector has emerged [2]. Currently, projector-camera systems use a grey level or RGB camera [3]. The main interest of color is to be able to differentiate geometrically similar patterns by color coding. Color coding also allows an easier matching of the points. Moreover, color data available in the acquired images may give the color of the object surface’s reconstructed points present in an object. However, this knowledge can be strongly skewed because of the limited number of color channels (three) within classical RGB camera. To overcome this limitation, we use a multispectral camera based on interference filters. A multispectral image is an image composed of several mono-channel images of the same object. In each image we have data about a specific wavelength according to the used interference filter. Such an imaging technique is becoming more and more interesting because of its great application potential [4].

The concept that we wish to describe consists in the use of the spectral reflectance as faithful representation of objects surfaces. To do this, we couple a multispectral camera with an LCD projector for scanning 3D artwork objects. In this case, a reflectance spectrum can be associated with each 3D reconstructed point. Knowing the spectral reflectance allows us to simulate the appearance of a 3D object under any virtual illuminant. Moreover, it allows storing this valuable information for future restoration.

The applications of such a system depend only on our imagination. Indeed, the system that we propose can also be useful for faithful digitalization of the objects of archaeology, digitalization of the cave paintings with relieves or digitalization of the walls of the caves with frescos. These objects are very fragile and will disappear one day, whatever the means invested to preserve them. The fact of digitizing them will preserve this memory, virtually at least. All these simulations find their interest in the case of virtual museum.

In Section 2, we first describe the experimental setup of the 3D Multispectral scanner and then we describe geometrical and spectral calibrations. Afterwards, we give the acquisition protocol in Section 3. In Section 4, we briefly treat the 3D geometrical reconstruction and then we explicit the spectral reconstruction. Some results related to the reconstruction of an artistic jug are given and discussed in Section 5. Section 6 is devoted to the illustration of some simulations enabled by the proposed system. It consists of illuminant change simulation and visualization of the 3D object in the framework of a virtual museum. Lastly, the conclusion completes this article.

2 System Description

2.1 Material

First of all, we developed a low cost multispectral imaging system [5]. It is designed to be portable and flexible, and is composed of a monochromatic CCD-based camera, a standard photographic lens, seven interference filters, a PC calculator, and C software developed especially for this system. A motorized wheel is placed in front of the camera/lens system (Fig. 1). The wheel has eight holes accepting seven filters and one empty hole, in order to make an acquisition without filter. Such a multispectral camera can reproduce the color with more precision than a traditional RGB camera since it is less affected by metamerism [6].

The complete 3D multispectral scanner system that we propose is composed of the camera detailed above and an LCD-projector. We chose an angle ranging between 35° and 40° between the camera sight axis and the LCD-projector one, which is the best compromise [7]. In addition to the luminosity, the depth-of-field and optical characteristics are important in the choice of an LCD-projector. A study that we carried out showed that an LCD-projector could be described by a pinhole type model [8]. In sum,
the results presented in this article are based on an object located at approximately 79” from the multispectral camera and LCD projector, with an area of 20” per 20”, and a depth of approximately 8”.

The complete 3D multispectral scanner system that we propose is composed of the camera detailed above and an LCD-projector. We chose an angle ranging between 35° and 40° between the camera sight axis and the LCD-projector one, which is the best compromise [7]. In addition to the luminosity, the depth-of-field and optical characteristics are important in the choice of the LCD-projector. A study that we carried out showed that an LCD-projector could be described using a pinhole type model [8]. In sum, the results presented in this article are based on an object located at approximately 79” from the multispectral camera and LCD-projector, with an area of 20” per 20”, and a depth of approximately 8”.

2.2 Calibration

Before any acquisition, spectral and geometrical calibrations are necessary. Once done, several acquisitions and reconstructions can be carried out without necessity to recompute these calibration parameters.

2.2.1 Geometrical calibration

The geometrical calibration is necessary to obtain 3D information. We chose a global solution which consists of calibrating the stereoscopic system set. This calibration method does not require any object of known size and can thus be carried out very easily. The camera is described by a pinhole model. Let \( M = (X \ Y \ Z \ 1)^T \) be homogeneous coordinates of a 3D-point in the reference frame of the object. Let \( m = (u \ v \ 1)^T \) be those of its projection in the image and expressed in pixels, then we can write:

\[
m = f(k, d, E, h, M)
\]  

where \( k \) is a vector of length 4 containing the intrinsic parameters: \((u0, v0)^T\) optical center coordinates and \( du, dv \), pixel size in the two directions. \( E \) size \([3 \times 3]\) and \( h \) size \([3 \times 1]\) are respectively 3D rotation, and 3D translation between the world and camera reference frames. \( d \) is the polynomial coefficients vector modeling the most significant distortions, the radial one. The total number parameters is 13. We showed that an LCD-projector could also follow this model [8]. Thus, the system calibration is similar to that of a standard stereoscopic system composed of two cameras. The only difference is based in the fact that the 3D characteristic points are not physically on
an object but emitted by the LCD-projector. Thus, we created an image pattern. It is made up of \( n \) luminous points on a dark background. This pattern is projected by the LCD-projector on a support with an unspecified position, and then it is acquired by the multispectral camera without filter. With the same pattern, this operation is repeated for \( q \) positions of this support. We need only that the luminous spots describe completely and regularly the work volume to be calibrated. With these conditions, Equation (1) can be written for the camera and for the LCD-projector. \( m_p \) is the point of the pattern projected on the support in \( M \) and \( m_c \) its projection in the image. We have 26 unknown factors for the whole system. The number of unknown factors is \( 26 + 3 \times n \times q \). Each function \( f \) corresponds to two equations. We obtain \( 4 \times n \times q \) equations. The system can thus be solved if it is overdetermined as it is the case for \( n \) and \( q \) are quite large. The establishment of the calibration parameters needs the minimization of the sum of the following equations:

\[
\| m_p - f(k_p, d_p, E_p, h_p, M) \|^2 + \| m_c - f(k_c, d_c, E_c, h_c, M) \|^2
\]  

for each \( n \times p \) points. This problem is nonlinear and we solved it by Levenberg–Marquardt optimization method [10].

2.2.2 Spectral calibration

With the spectral model of the acquisition chain (Fig. 2), the signal \( d_k \) observed from the camera output, relative to channel \( k \) \((k=1...7)\), is given by Equation (3):

\[
d_k = \int_{\lambda_{\text{min}}}^{\lambda_{\text{max}}} I(\lambda) r(\lambda) c(\lambda) t_k(\lambda) o(\lambda) d\lambda + \eta_k
\]

where \( I(\lambda) \) is the spectral radiance of the illuminant, \( r(\lambda) \) is the spectral reflectance of the surface, \( c(\lambda) \) is the camera spectral sensitivity, \( t_k(\lambda) \) is spectral transmittance according to the filter number \( k \), \( o(\lambda) \) is the transfer function of optics, and \( \eta_k \) expresses the spectral noise of the \( k^{th} \) filter.

Fig. 2. Synopsis of the spectral model of the acquisition process in a multispectral system.

The term reflectance contains in general two kinds of reflections: diffuse and specular. From the model above, we seek to recover diffuse reflection. So, it is necessary to estimate and remove specular reflection. In literature, there are three kinds of methods to separate diffuse and specular components [11]:

– first, methods using polarizing filters such as [12] [13],
– secondly, methods using more than one image [14] [15],
– finally, methods using color and requiring only a single input image [16] [17].

These latter are relatively more practical in general conditions but most of them require color segmentation to deal with multicolored images. In our side, we take care to use a diffusing light with appropriate angles during the experiments. We thusly manage to reduce specular reflection. Afterwards, in an image processing step, we detect specular pixels by an adaptive thresholding. The specular pixels are those pixels that are both greater than the mean of the regions and also that of the whole image. Then, they are fulfilled in with the spline interpolated values calculated using the nearest legitimate pixels. In doing so, we significantly reduce the effect of the specular reflection.

In order to faithfully attain sought-after data during acquisitions, namely the object spectral reflectance, a set of radiometric calibrations must be done. The first one is the treatment of the spectral noise, called thusly because it is specific to each channel. The interested reader can refer to the following article [9]. Preprocessing and noise reduction finished, we can spectrally characterize our system. The goal is to determine the spectral sensitivity of the system for each channel according to the spectral model of Equation (3). This spectral model is based on a linear opto-electronic transfer function assumption. This assumption generally holds when noise is reduced and makes Equation (3) as a simple multiplication of spectra contained wavelength by wavelength in vectors. So, the spectral sensitivity of the system of the channel \( k \), \( S_k(\lambda) \), contains the radiance of illumination \( I(\lambda) \), the sensitivity of the sensor \( c(\lambda) \), the response of optics \( o(\lambda) \), and the transmittance of the filter \( t_k(\lambda) \). By sampling out of \( N \) regular intervals the spectrum range in which we work, we can rewrite Equation (3) in matrix notation. In our case, the spectrum range extends from \( \lambda_1 = 380 \text{nm} \) (nanometers) to \( \lambda_N = 780 \text{nm} \) and the sampling step is \( 5 \text{nm} \) which gives a value of 80 for \( N \). This interval also corresponds to the sensitivity range of the spectrophotometer used. Equation (3) becomes:

\[
d_k = r(\lambda)^T S_k(\lambda)
\]

where \( S_k(\lambda) = [S_k(\lambda_1), S_k(\lambda_2), \ldots, S_k(\lambda_N)]^T \) and \( r(\lambda) = [r(\lambda_1), r(\lambda_2), \ldots, r(\lambda_N)]^T \) are respectively the vectors containing the spectral sensitivity of the acquisition system relating to the channel \( k \), and spectral reflectances. \( ^T \) is the transposed matrix operator.

3 Acquisition

The two calibration stages finished (geometrical and spectral), it is now possible to acquire as many objects as we wish as long as the acquisition configuration remains unchanged. It suffices to put the object to be reconstructed in the calibrated work volume. In order to highlight the application for 3D artwork objects, we used an artistic jug for our acquisitions. First, a multispectral image of the object is acquired without light projection by the LCD-projector. This image will thereafter allow, during the spectral reconstruction detailed in paragraph 4.2.2., to correlate a reflectance spectrum to each reconstructed 3D-point of the object. Then, a set of images is acquired without filter. For each one, the LCD-projector emits a sufficiently intense luminous vertical line so that it appears on the surface of the object and thus on the images. The line is shifted pixel by pixel between each image. We repeated these process each time we rotate the jug. In this circumstances, we rotated the object three times in order to get a sight of \( 180^\circ \). The following Section describe the necessary treatments for reconstructing the 3D-object in both geometrical and spectral aspects.

4 Reconstruction

4.1 Geometrical reconstruction

The camera and the LCD-projector being placed at the same height and approximately at the same distance from the object, a vertical line emitted by the LCD-projector
sweeps the object. Because the distortion parameters of the LCD-projector are weak, we supposed them negligible. That is why 2D line in the LCD-projector image plane describes, in space, a 3D plane. In camera images, we can detect luminous pixels. This point and the camera optical center are on a line of sight. We can finally calculate the intersection of the luminous plane and the line of sight. Geometrical reconstruction allows us to obtain the 3D position of the different parts of the luminous pattern on the object by using triangulation.

4.2 Spectral reconstruction

4.2.1 Problem Position According to the Equation (4), we now address the problem of how to estimate the spectral reflectance $r$ of an object from the camera responses $d_k$. This is an inverse problem. So, finding appropriate mathematical methods to perform this inversion is of great importance. Several methods to achieve this task exist in literature. Some classical approaches [18][19][20] use the pseudo-inverse calculus and the least squares. The main drawback of these methods is the noise amplification. Some other methods seek to maximize the smoothness of the estimate result [21][22]. Another reliable method proposed in [4] takes advantage of the a priori knowledge on the spectral reflectances of surfaces that are to be imaged. In our side, we use a learning-based method to invert the Equation (2) based upon neural networks: associative memories [23]. This method presents two advantages: first it is robust to noise affecting input data and secondly it is based on combination of several neural networks. That allows good generalization and hence enables reconstruction of a wide range of reflectances, even those for which the memories were not trained.

From Equation (4), we first search to characterize the spectral response of the system by finding the operator $S = [S_1(\lambda) S_2(\lambda) ... S_7(\lambda)]^T$. The $S$ matrix size is $[7 \times 80]$ and represents the spectral response of the complete system with all 7 channels. To do this, we scanned the Macbeth chart using a MinoltaCS−1000 spectrophotometer and we acquired a multispectral image of this chart. The result is a set of corresponding pairs $(d_p, r_p)$, for $p=1,...,24$, where $d_p$ is a vector of dimension $k=7$ containing the camera output and $r_p$ is a vector of dimension $N$ representing the spectral reflectances of the $p^{th}$ patch. By observing the camera output responses, we can estimate the system response to known theoretical reflectances in the input. This stage is called learning because we use a neural network to invert the Equation (4). Specifically, we use a linear neural network associative memories [23]. The use of linear operator is justified by the fact that we supposed a linear opto-electronic transfer function and since the noise have been reduced in the pre-processing stage.

4.2.2 Reconstruction using neural network Artificial neural networks simulate the behavior of many simple processing elements present in the human brain, called neurons. Each neuron is linked to each other by connections called synapses. Each synapse has a coefficient that represents the strength or weight of the connection. Learning or training is accomplished by adjusting these weights to cause the neural network to appropriate output results. The learning rules specify how to calculate the modification of the weights based on the objective. The perceptron (Fig. 3) is the first and basic model of existing neural networks.

The symbol $a_j$ expresses the activation of the $j^{th}$ output cell, $x_i$ is the answer of the $i^{th}$ input cell and $w_{ij}$ is the intensity connection between $i^{th}$ input cell and $j^{th}$ output cell. The answer $o_j$ of the $j^{th}$ output cell, which is function of the neuron activation $a_j$, is only 0 or 1 in the case of deterministic perceptron. Because the perceptron gives the same response to the same stimulus after training, we modify its behavior by modifying the function $f$ that translates the activation into a response. In our case, we use the Boltzmann distribution. The perceptron gives now probabilistic response. This option corresponds to the creation of associative memories. Associative memories
can be hetero-associative or auto-associative. Hetero-associative memories associate a stimulus in the input to a response in the output even if the two vectors have not the same size, where auto-associative memories (particular case) associate a stimulus to itself and can, thusly be used to store stimuli. Associative memories meet our needs because we want to associate an input vector corresponding to the 7 gray-level values of a pixel to an output vector representing the spectral reflectance sampled with \( N = 80 \) values. From Equation (2), we can see the problem like that: we search a set of values of \( w_{ij} \) such that the memory associates a configuration on the \( N \) output cells with stimuli presented in the \( k \) input cells (corresponding to \( k \) channels) of the neural network. The memory associates a vector of 7 values coming from the multispectral image to a vector of \( N=80 \) values. In matrix notation, we express this as finding a matrix of weight \( W \) of order \( N \times k \) with:

\[
e_p = o_p = W^T x_p,
\]

where \( e_p \) is the vector containing the expected response, \( x_p \) is the vector containing the \( k \) input values coming from the multispectral image, and \( o_p \) is the output for the stimulus \( P \). Finding \( W \) corresponds to learning and leads to determine the system response \( S \) for all the \( k \) channels.

First, determining \( W \) corresponds to determining the system response \( S \) for all the 7 channels. The second step is the reconstruction. As for the geometrical characterization, it is necessary to preserve the camera acquisition parameters and not to modify the illuminant.

The spectral characteristics \( S \) of the multispectral system are henceforth known. The estimate of a reflectance spectrum \( \tilde{r} \) in each pixel of an acquired object with this system is thus possible. The vector \( d = [d_1, d_2, ..., d_7]^T \) containing the responses for the 7 filters is given by Equation (4). The second step is the reconstruction. Since all weighted synapses are gathered in the matrix \( S \), the reconstruction is fast and easy: the estimated spectral reflectance in each pixel of a scene is equal to a product between the operator \( S \) and the camera response contained in \( d \).

During the training, the memory may learn only few samples among possible stimuli. The memory stops learning when it ceases to mistake any more. So, it may place the discriminating function too much close to the boundaries of the samples with which it was trained. If we test it on new samples, the memory may badly generalize its training. To overcome this problem, we use several associative memories in cascade schemes with 'Delta' as a rule of training which. It consists on continuously modifying the strengths of the weight matrix. In doing so, the memory becomes more efficient for generalization. Indeed, this algorithm allows to reconstruct spectra which were not learned. This result comes from the fact that the associative memories are based on Principal Component Analysis (PCA).

5 Results and Discussion

In order to evaluate the 3D reconstruction error, we compared the results of our scanning method with those coming from the Minolta VIVID – 910 professional scanner.
The Fig. 4.a illustrates the cloud of points reconstructed with our method. Although our geometrical reconstruction gives a weak number of points, we present a triangulated surface from the 3D-points (Fig. 4.b). The Fig. 4.c is the result obtained from the professional scanner. The comparison of the surfaces from the two systems shows that 90% of the values are between $-0.03''$ and $0.03''$ (Fig. 5). On this figure, we show the distance card between the reconstructed cloud and one issued from the professional scanner for the three position of the jug. It is clearly seen that the strongest errors are on the edges of the object and in area with strong curvatures. The scanning method that we use involves such errors because we have a weak in-depth precision when scanning away from the camera axis.

![Fig. 4. Geometrical reconstruction; a. cloud of points obtained by 3D reconstruction, b. Triangulated surface, c. Surface obtained with the professional scanner.](image)

The use of the multispectral camera enables us to go up with the spectral reflectance of object surfaces. This information can be attached to each 3D reconstructed point. It is valuable information because it represents a physical property which depends neither on the illuminant nor on the subjectivity of human vision.

![Fig. 5. Distance card between the reconstructed cloud and one issued from the professional scanner for the three position of the jug.](image)

In order to validate the reconstruction of the spectral reflectance by the suggested method, we made a comparison, for a set of pixels, between the measured spectrum using the spectrophotometer and the one reconstructed from the response of the multispectral camera. We used the the Goodness-of-Fit Coefficient \((GFC)\) as criterion [24].
given by the following formula:

\[ GFC = \frac{\left| \sum_j R_m(\lambda_j) R_r(\lambda_j) \right|}{\left( \sum_j [R_m(\lambda_j)]^2 \right)^{1/2} \left( \sum_j [R_r(\lambda_j)]^2 \right)^{1/2}} \]  

(6)

where \( R_m(\lambda_j) \) is the value of the measured spectrum with the spectrophotometer at the wavelength \( \lambda_j \), and \( R_r(\lambda_j) \) represents that related to the reconstructed spectrum at the wavelength \( \lambda_j \). Within the meaning of this criterion, the results are very satisfactory. The reconstructed spectrum shows few errors compared to that obtained by the spectrophotometer. We note an average error of the \( GFC \) equal to 1.3% for a standard deviation of 0.1%.

6 Simulations

Once we have the reconstructed 3D object, each 3D point is associated with a spectral reflectance. So, we have a 3D spectral object. In order to visualize it, we choose an illuminant and associate a chromatic RGB triplet to each point. Note that we are able to simulate and visualize the object such as it will be with any illuminant. In order to emphasize this point, the Fig. 6 presents a 3D spectral jug visualized with two different illuminants and finally projected in RGB space. Let us notice that the left and right images come from the same 3D spectral jug.

![Fig. 6. Projection of the 3D spectral object in RGB color space after simulation of illuminant change. a. object appearance under the CIE A illuminant, and b. object appearance under the CIE D65 illuminant (daylight).](image)

The Fig. 7, illustrates some examples of the application of virtual museum. This concept provides a real opportunity for museums to virtually and permanently exhibit their artworks, and for visitors a real flexibility in that sense it is accessible via internet or via a DVD-ROM. Like the example in the Fig.7, a virtual visitor browses between several thematic rooms where artworks are exposed. When the visualized artwork is 2D one (e.g Van-Gogh autoportrait), the visitor can simulate illuminant changes and
listen comments about the object that he is visualizing. When the objects are three-dimensional artwork or archeological finds, the visitor can listen comments about these objects and handle them easily and securely by simple mouse clicks (e.g. the 3D spectral jug can be rotated spatially and interactively visualized under any illuminant in the web browser). In this context, we can cite some interesting works of Zheng et al. [25] [26]. In their study, the authors use an active system based on a single camera and a laser projector for the aim of virtual recovery of excavated archaeological finds. They propose a system which enables accurate 3D shape reconstruction and the color information is extracted from remaining color samples to the object surfaces where the ancient pigments have dropped. In our side, we seek to recover, instead of color, the spectral reflectance which is a physical representation of scanned surfaces. This representation is independent from both the acquisition system and the observer and permits faithful archiving and interesting multimedia applications.

![Fig. 7. Examples of the simulations enabled by the proposed system in the framework of a virtual museum.](image-url)
7 Conclusion

We just presented an active stereovision system based on a LCD-projector and a camera. The characteristic of this camera is that it is a multispectral one. We aim at the development of a multispectral scanner for 3D artwork objects for multimedia applications. The contribution of the multispectral concept comes back to the possibility of reconstructing and associating a reflectance spectrum for each 3D-point. Thus, it produces much more relevant information than during the use of conventional color camera. The results presented proved the feasibility of such a system that we named "3D multispectral scanner". Measured errors at the geometrical and spectral level, remain relatively weak. Our current work consists of generating and emitting more complex luminous patterns on the object. The goal is to decrease the number of acquisitions while preserving a dense object reconstruction.

Acknowledgements

This work was supported by CNRS, Burgundy region, and FEDER.

References


