A stereoscopic system based on a multispectral camera and an LCD projector uses multiplex spectral information for 3D object reconstruction. By linking 3D points to a curve representing the spectral reflectance, the system gives a physical representation of the matter that's independent from illuminant, observer, and acquisition devices.

Vision-based systems are becoming increasingly prevalent in the 3D object reconstruction and metrology fields. In particular, these systems seem preferable to techniques based on interferometry or the watered effect within the framework of applications whose objects have significant volume, because the latter mainly measure small objects in nanometer-order depths.

Vision systems can be passive or active. Passive systems acquire a 3D object using cameras only. They can use one or several cameras to acquire the same object, but they must always acquire at least two images of the object from different view angles. They give position and depth information by matching pixels between images to reconstruct by triangulation their 3D shape. In such methods, low-textured objects become difficult to analyze because few characteristic points appear on the surface.

Active vision systems use a device that projects a structured light onto the object. This light can be a laser beam or a light from a liquid crystal display (LCD) projector. The structured light creates a texture on the object’s surface that a camera can then acquire. If the system is geometrically calibrated, it’s easy to calculate the object points’ positions and depth, as illuminated by the projected pattern. Researchers have studied many types of structured light; Battle, Mouaddib, and Salvi describe some state-of-the-art techniques.

Within this framework, systems consisting of a camera and an LCD projector have emerged. Current projector-camera systems use a gray-level or red-green-blue (RGB) camera. They typically use color to differentiate geometrically similar patterns. Color coding also facilitates point matching. Moreover, color data from the acquired images can give the color of the object surface's reconstructed points. However, the limited number of color channels (three) in classical RGB cameras can skew this knowledge. To overcome this limitation, we use a multispectral camera based on interference filters. A multispectral image is composed of several monochannel images of the same object. Each image holds data about a specific wavelength depending on the interference filter used. This imaging technique has great application potential.

To produce a spectral reflectance that can faithfully represent an object surface, we couple a multispectral camera with an LCD projector for scanning 3D artwork objects. We associate a reflectance spectrum with each 3D reconstructed point. Knowing the spectral reflectance lets us simulate the appearance of a 3D object under any virtual illuminant. Moreover, it lets us store this valuable information for future restoration.

System description

We designed our low-cost multispectral imaging system to be portable and flexible. The system consists of a monochromatic charge-coupled device (CCD)-based camera, a standard photographic lens, seven interference filters, a PC calculator, and custom C software. As Figure 1 shows, we place a motorized wheel in front of the camera/lens system. The wheel has eight holes holding seven filters and one empty hole for acquiring images without a filter. Such a multispectral camera can reproduce color with more precision than a traditional RGB camera because it’s less affected by metamerism—that is, being split into homologous segments.

The complete 3D multispectral scanner system consists of the camera and an LCD projector. In a previous study, we showed that the optimal angle between the camera sight axis and the LCD projector is about 35 to 40 degrees. In addition to luminosity, important considerations in choosing an LCD projector are depth-of-field and optical characteristics. An earlier study showed that
we can describe an LCD projector by its pinhole model. We based the results presented in this article on an object located approximately 79 inches from the multispectral camera and LCD projector, with an area of 20 inches by 20 inches, and a depth of approximately 8 inches.

**Calibration**

Before we can acquire images, we must perform spectral and geometrical calibrations. Once this is completed, the system can perform several acquisitions and reconstructions without our needing to recompute the parameters.

**Geometrical calibration**

Geometrical calibration is necessary to obtain 3D information. We chose a global solution that consists of calibrating the stereoscopic system set. This calibration method doesn’t require an object of known size and is thus easy to perform. We use a pinhole model to describe the camera. Let \( M = (X \ Y \ Z \ 1)^T \) be homogeneous coordinates of a 3D point in the object’s reference frame. Let \( m = (u \ v \ 1)^T \) be the coordinates of the object’s projection in the image expressed in pixels. Then,

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

where \( k \) is a vector of length 4 containing the following intrinsic parameters:

- \((u_0, v_0)^T\) are the optical center coordinates, and
- \(du, dv\) is the pixel size in the two directions.

The operators \( E \) (size \( 3 \times 3 \)), and \( h \) (size \( 3 \times 1 \)), are 3D rotation and 3D translation between the world and camera reference frames, respectively. The polynomial coefficients’ vector \( d \) models the radial distortion (the most significant distortion). We use a total of 13 parameters.

Because, as we showed in previous work, an LCD projector can follow this model, the system calibration is similar to that of a standard two-camera stereoscopic system. The only difference is that the 3D characteristic points aren’t physically on an object but emitted by the LCD projector. We therefore created an image pattern made up of \( n \) luminous points on a dark background. The LCD projector casts this pattern on a support with an unspecified position. The multispectral camera without a filter then captures the pattern.

Using the same pattern, we repeat this operation for \( q \) positions of the support. We need only that the luminous spots describe completely and regularly the work volume to be calibrated. With these conditions, we can write Equation 1 for the camera and LCD projector where \( m_p \) is the point of the pattern projected on the support in \( M \) and \( m_c \) its projection in the image.

The entire system is described by 26 parameters. So, the number of unknown factors for the \( n \times q \) emitted positions is \( 26 + 3 \times n \times q \). Each function \( f \) corresponds to two equations. We obtain \( 4 \times n \times q \) equations. We can thus solve the system if it’s overdetermined, as is the case when \( n \) and \( q \) are quite large. To establish the calibration parameters, we minimize the sum of Equation 2:

\[
\sum \|m_p - f_p(k_p, d_p, E_p, h_p, M_p)\|^2 + \|m_c - f_c(k_c, d_c, E_c, h_c, M_c)\|^2
\]

for each \( n \times p \) points.

**Spectral calibration**

Equation 3 describes the signal \( d_k \) observed from the camera output relative to channel \( k \) (\( k = 1 \ldots 7 \)) using the spectral model of the acquisition chain (see Figure 2):

![Figure 1. Our multispectral system: (a) internal view and (b) external view.](image1)

![Figure 2. The spectral model of the acquisition process in a multispectral system.](image2)
mittance according to the filter number
the optics' transfer function, and
spectral sensitivity of channel
3 a simple multiplication of spectra contained
holds when noise is reduced and makes Equation
function assumption. This assumption generally
model is based on a linear optoelectronic transfer
the spectral model in Equation 3. This spectral
spectral sensitivity for each channel according to
our system. The goal is to determine the system's
noise reduction, we can spectrally characterize
the radiance of illumination
derature describes three methods for separating
estimate and remove specular reflection. The lit-
must recover diffuse reflection, we must
reflection: diffuse and specular. Because our
camera spectral sensitivity,
r
I
k
acquisition
nificantly reduce the specular reflection's effect.
in the nearest legitimate pixels. In doing so, we sig-
spline-interpolated values that we calculate using
both greater than the mean of the regions and
olding. Specular pixels are those pixels that are
horizontal, pixel by pixel between each image,
iciently intense to appear on the object's surface
let us correlate a reflectance spectrum to each of
object without light projection. This image will
object's reconstructed 3D points during spec-
jects to put the object to be reconstructed in the
reasonable to each channel.
range from
Equation 3 becomes:
\[ d_k = r(\lambda)S_k(\lambda) \] (4)
where \( S_k(\lambda) = [S_1(\lambda_1) \ S_2(\lambda_2) \ldots S_N(\lambda_N)]^T \) and
and \( r(\lambda) = [r(\lambda_1) \ r(\lambda_2) \ldots r(\lambda_N)]^T \) are, respectively, the vectors
containing the spectral sensitivity of the acquisi-
tion system relating to channel \( k \) and spectral
reflectances. \( T \) is the transposed matrix operator.

Acquisition
Once calibration is complete, we can acquire
as many objects as we wish, as long as the acqui-
sition configuration remains unchanged. It suf-
fices to put the object to be reconstructed in the
 calibrated work volume. To highlight the appli-
cation for 3D artwork objects, we used an artistic
jug for our acquisitions.
First, we acquire a multispectral image of the
object without light projection. This image will
let us correlate a reflectance spectrum to each of
the object's reconstructed 3D points during spec-
tral reconstruction. Next, we acquire a set of
images without a filter. For each image, the LCD
projector emits a luminous vertical line suffi-
ciently intense to appear on the object’s surface
and thus on the images. We shift the line hori-
zontally, pixel by pixel between each image,
repeating these processes each time we rotate the
jug. In this case, we rotated the object three times
to get a 180-degree sight.

Reconstruction
Reconstructing the 3D object requires geo-
matical and spectral treatments.

Geometrical reconstruction
We place the camera and the LCD projector
at the same height and approximately the same
distance from the object. The LCD projector
emits a vertical line that sweeps the object.
Because the LCD projector’s distortion param-
ters are weak, we assume they're negligible,
which is why a 2D line in the LCD projector
image plane describes, in space, a 3D plane. In
camera images, we can detect luminous pixels. Each pixel describes a 3D point. Because this point and the camera optical center are on a line of sight, we can calculate the intersection of the luminous plane and the line of sight. Geometrical reconstruction lets us obtain the 3D position of the different parts of the luminous pattern on the object using triangulation.

**Spectral reconstruction**

We now address the inverse problem of estimating an object’s spectral reflectance \( \mathbf{r} \) from the camera responses \( d_i \) (see Equation 4). It’s important to find appropriate mathematical methods to perform this inversion. Several methods to achieve this task exist in the literature. Some classical approaches use the pseudo-inverse calculus and least squares. The main drawback of these methods is the noise amplification. Some other methods seek to maximize the estimated result’s smoothness.\(^3\) Another reliable method exploits the a priori knowledge of the surfaces’ spectral reflectances to be imaged.\(^3\)

We use associative memories, a learning-based method, to invert Equation 2 based on neural networks. This method has two advantages:

- It’s robust to noise affecting input data.
- It’s based on a combination of several neural networks.

These advantages allow good generalization and hence let us reconstruct a wide range of reflectances, even those for which the memories weren’t trained.

From Equation 4, we first characterize the spectral response of the system by finding the operator \( \mathbf{S} = [S_1(\lambda) S_2(\lambda) \ldots S_7(\lambda)]^T \). The \( \mathbf{S} \) matrix size is \( 7 \times 80 \) and represents the complete system’s spectral response with all seven channels. We scan the Macbeth chart using a Minolta CS 1,000 spectrophotometer to acquire a multispectral image of the chart. The result is a set of corresponding pairs \( (d_p, r_p) \), for \( p = 1, \ldots, 24 \), 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. We call this the learning stage because we use a neural network to invert Equation 4. By assuming that it’s a linear opto-electronic transfer function and by reducing the noise in the preprocessing stage we can justify using a linear operator.

Artificial neural networks simulate the behavior of neurons, the simple processing elements present in the human brain. Synapses link neurons together. Each synapse has a coefficient representing the connection’s strength or weight. Learning or training is accomplished by adjusting these weights so the neural network produces appropriate results. Learning rules specify how to calculate the weight modifications based on the objective. The perceptron (Figure 3) is the first and basic model of existing neural networks. The symbol \( a_j \) expresses the \( j \)th output cell’s activation, \( x_i \) is the \( i \)th input cell’s answer, and \( W_{ij} \) is the intensity connection between the \( i \)th input cell and the \( j \)th output cell. The answer \( o_j \) of the \( j \)th output cell, which is a function of the neuron activation \( a_j \), is only 0 or 1 in the case of a 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.

Our system uses the Boltzmann distribution. The perceptron now gives a probabilistic response. This option corresponds to the creation of associative memories, which can be either

- **hetero-associative** memories, which associate a stimulus in the input to a response in the output even if the two vectors aren’t the same size, or
- **auto-associative** memories, which associate a stimulus to itself and can therefore be used to store stimuli.

Associative memories meet our needs because we want to associate an input vector corresponding to a pixel’s seven gray-level values to an output vector representing the spectral reflectance sampled with \( N = 80 \) values. From Equation 2, we can see the problem as searching 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.

![Figure 3. One output cell of a perception.](image-url)
The memory associates a vector of seven values coming from the multispectral image to a vector of \( N / H^2 \) values. In matrix notation, we express this as finding a matrix of weight \( W \) of order \( N / H^0 \) 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 from the multispectral image, and \( o_p \) is the output for the stimulus \( P \). Finding \( W \) corresponds to learning and lets us determine the system response \( S \) for all \( k \) channels.

Once we know the multispectral system’s spectral characteristics \( S \), we can estimate a reflectance spectrum in each pixel of an acquired object. Equation 4 gives vector \( d = [d_1, d_2, ..., d_7]^T \) containing the responses for the seven filters.

Reconstruction is the second step. Because all weighted synapses are gathered in the matrix \( S \), 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 might learn only a few of the possible stimuli. The memory stops learning when it makes no more mistakes. So, it might place the discriminating function too close to the boundaries of the samples with which it was trained. If we test it on new samples, the memory might badly generalize its training. To overcome this problem, we use several associative memories in cascade schemes with “Delta” as a training rule. This rule consists of continuously modifying the weight matrix’s strengths, which makes the memory more efficient for generalization. Indeed, this algorithm lets us reconstruct spectra that weren’t learned, a result of our basing the associative memories on principal component analysis.

Results

To evaluate the 3D reconstruction error, we compared the results of our scanning method with those of the Minolta Vivid 910 professional scanner. Figure 4a shows the point cloud reconstructed with our method. Although our geometrical reconstruction gives a weak number of points, we present a triangulated surface from the 3D points (see Figure 4b). Figure 4c shows the results from the professional scanner.

As Figure 5 shows, a comparison of the surfaces from the two systems shows that 90 percent of the values are between ~0.03 inches and 0.03 inches. The figure shows the distance card between the reconstructed cloud and a cloud issued by the professional scanner for the jug’s three positions. The strongest errors are on the object’s edges and in areas with strong curvatures. Our scanning method involves such errors because we have a weak in-depth precision when scanning away from the camera axis. Using the multispectral camera lets us recover the spectral reflectance of object surfaces. We can attach this information to each 3D-reconstructed point. The information is valuable because it represents a physical property that depends neither on the illuminant nor on human vision’s subjectivity.
To validate the reconstruction of the spectral reflectance by the suggested method, we compared, for a set of pixels, the measured spectrum using the spectrophotometer and the spectrum reconstructed from the multispectral camera’s response. We used the goodness-of-fit coefficient (GFC), given by the following formula, as criterion:

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

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

**Simulations**

After reconstructing the 3D object, we associate each 3D point with a spectral reflectance, giving us a 3D spectral object. To visualize the object, we choose an illuminant and associate a chromatic RGB triplet to each point. We can simulate and visualize the object as it can be seen under any illuminant. To emphasize this point, Figure 6 presents a 3D spectral jug visualized with two different illuminants projected in RGB space. Notice that the left and right images are from the same 3D spectral jug.

Figure 7 shows some examples of the virtual museum application. Such an application lets museums virtually and permanently exhibit their artwork, and gives visitors the flexibility to access the artwork via the Internet or a DVD-ROM. As Figure 7 illustrates, a virtual visitor can browse several thematic rooms containing artwork. When the visualized object is 2D (such as a van Gogh autoportrait), the visitor can simulate illuminant changes and listen to comments about the object being visualized. When the objects are 3D artwork or archaeological finds, the visitor can listen to comments about them and handle them easily and securely using simple mouse clicks (for example, the user can rotate the 3D spectral jug spatially and visualize it interactively under any illuminant in the Web browser).

Other researchers have done some interesting work in this area. For example, Zheng uses an...
active system based on a single camera and a laser projector for virtual recovery of excavated archaeological finds. Their system enables accurate 3D shape reconstruction and transfer of color information from remaining color samples to the object surfaces where the ancient pigments have dropped. Instead of color, our work seeks to recover the spectral reflectance, a physical representation of the scanned surfaces. This representation is independent of both the acquisition system and the observer and permits faithful archiving as well as interesting multimedia applications.

Conclusion
The applications of such a system depend only on our imagination. For example, we could use the system to digitize archaeological objects, such as cave paintings, letting us preserve these fragile objects virtually, if not in reality.

Current work consists of simultaneously generating and emitting several luminous patterns on the scene. The goal is to decrease the number of acquisitions while preserving a dense scene reconstruction.

Acknowledgments
The French National Council for Scientific Research, Burgundy region, and FEDER (Fonds Européens de développement Régional) supported this work.

References

Alamin Mansouri is a postdoctoral researcher at i3mainz Lab, Germany, where he’s developing and implementing algorithms for photogrammetry using both colorimetric and multispectral approaches. His research interests include color and multispectral imaging and 3D reconstruction. Mansouri has a PhD in computer science from the Laboratory of Electronics, Information, and Imaging (Le2i), University of Burgundy, France.
Alexandra Lathuilière is a graduate student in the Le2i in Le Creusot, France. Her research interests include the automatic adaptation of color-structured light patterns for 3D scene reconstruction for active stereovision systems. Lathuilière has a DEA degree in images from the Institut Supérieur des Techniques Avancées at the University of Saint Etienne, France (ISTASE).

Franck S. Marzani is an associate professor of computer science at the University of Burgundy and a member of the Le2i. His research interests include multispectral images and range-sensing systems by structured light. Marzani has a PhD in computer science from the University of Burgundy, Dijon, France.

Yvon Voisin is a full professor of signal and image processing at the University of Burgundy and a member of the Image Processing Group at the Le2i. His research interests are 3D reconstruction and motion analysis. He’s also working on the application of artificial vision, especially in biology. Voisin has a PhD in electronic and signal processing from the University of Franche Comté, France.

Pierre Gouton is a professor of electronic and signal processing at the University of Burgundy and manager of a research team within the Le2i. His research interests are color and multispectral imaging, 3D imaging, and remote sensing. Gouton has a PhD from the University of Montpellier, France.

Readers may contact Alamin Mansouri at Le2i, UMR CNRS 5158, Université of Burgundy, Aile des sciences de l’ingénieur, BP 47870 21078 Dijon CEDEX, France; alamin.mansouri@u-bourgogne.fr.

For further information on this or any other computing topic, please visit our Digital Library at http://computer.org/publications/dlib.