Intrinsic Images for Dense Stereo Matching with Occlusions

César Silva and José Santos-Victor

Instituto de Sistemas e Robótica
Instituto Superior Técnico
Av. Rovisco Pais, 1, 1049-001 Lisboa, Portugal
{cesar,jasv}@isr.ist.utl.pt

Abstract. Stereo correspondence is a central issue in computer vision. The traditional approach involves extracting image features, establishing correspondences based on photometric and geometric criteria and finally, determine a dense disparity field by interpolation. In this context, occlusions are considered as undesirable artifacts and often ignored.
The challenging problems addressed in this paper are a) finding an image representation that facilitates (or even trivializes) the matching procedure and, b) detecting and including occlusion points in such representation.
We propose a new image representation called *Intrinsic Images* that can be used to solve correspondence problems within a natural and intuitive framework. Intrinsic images combine photometric and geometric descriptors of a stereo image pair. We extend this framework to deal with occlusions and brightness changes between two views.
We show that this new representation greatly simplifies the computation of dense disparity maps and the synthesis of novel views of a given scene, obtained directly from this image representation. Results are shown to illustrate the performance of the proposed methodology, under perspective effects and in the presence of occlusions.

1 Introduction

Finding correspondences is one of central problems in stereo and motion analysis. To perceive motion and depth, the human binocular vision uses shape and edge segmentation and performs matching between two or more images, captured over time. Besides, occlusions play a key role in motion and depth interpretation.

The computation of stereo correspondence has traditionally been associated with a generic three-step procedure [4], summarized as follows:

- Selecting image features, such as edges, interest points or brightness values.
- Find corresponding features based on similarity and consistency criteria, where similarity considers a distance between features and consistency takes into account geometric and order constraints. This step is usually algorithmic and requires a relative weighting between similarity and consistency.
- Compute and interpolate the disparity maps to obtain a dense field.

An additional step consists of detecting occlusion points, where the similarity and/or consistency criteria are not fulfilled.

Almost all methods for image correspondence follow the procedure described above and differ only in the nature of the image features, similarity and consistency criteria, search algorithms and photometric and geometric assumptions [5, 9, 8] (see [4] for an overview).

We propose a method where the global stereo image information and the similarity and consistency criteria are well defined in a common, natural and intuitive framework. This approach can be used to generate a dense disparity map (explicit reconstruction) or different views of the same scene (without explicit reconstruction).

1.1 Motivation

One of the challenges for solving the correspondence problem is that of finding an image representation that facilitates (or even trivializes) the matching procedure. As an example, consider two corresponding epipolar lines of the stereo image pair shown in Figure 1. The simplest function we can analyse is the brightness function $f(y)$ and $g(x)$ defined along each (left or right) epipolar line — Figure 1a. This figure shows the difficulty of the matching process since the gray level functions along both epipolar lines are geometrically deformed by the 3D structure. However, we can obtain other representations for the information included along a scanline, as follows:

1. A commonly used representation, mainly in optical flow computation [6], consists in the spatial derivatives of the brightness (Figure 1b). Yet the matching process is not trivial specially for a wide baseline stereo system.
2. To search for correspondences, one could use the plot of Figure 1a and determine the points with equal brightness value. However, this would only work when the brightness is kept exactly constant and would lead to many ambiguities, as illustrated in the figure for the gray value 240.
3. Other possible representation consists in plotting the brightness versus its derivative [8] as shown in Figure 1c. In this case, the image points with the same brightness and derivatives have approximately the same coordinates, indicating a correspondence. Again, there are some ambiguous situations (shown in the figure) and the points are coincident only if the disparity is constant (no perspective effects) along these image lines.

These representations can be generalized considering other local descriptors (other than brightness, position and spatial gradient) computed along two or more corresponding scanlines. Tomasi and Manduchi [8] have proposed to represent a set of local descriptor vectors (brightness, derivative, second derivative, etc) through a curve in a n-dimensional space, where the curve represented in Figure 1c is a simple example. Ideally two curves computed along two corresponding scanlines can be mapped (or even coincide).

However, approaches based on curves of local descriptors vectors have obvious limitations related to rigid geometric distortion assumptions, solution ambiguity

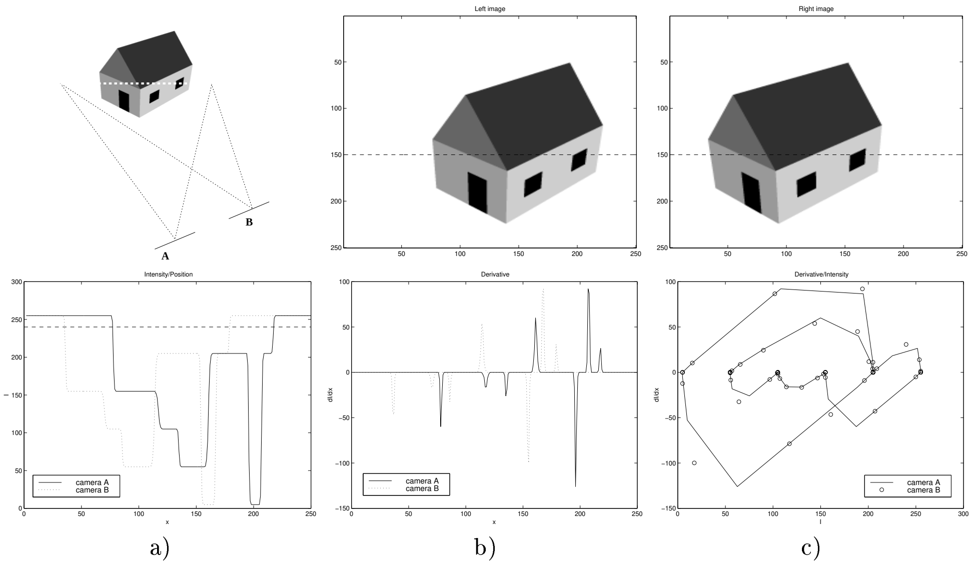


Fig. 1. Photometric and geometric relations of the brightness values along a scanline captured by a stereo pair. a) Brightness values versus pixel positions; b) Derivatives of the brightness versus pixel positions; c) Derivatives versus brightness values.

and/or high-dimensionality search algorithms. First of all, the method is only valid for constant and affine disparities (no perspective effects have been considered). Secondly, the curves have a difficult representation, specially if more than two local descriptors are considered. Finally, a curve can cross itself, clearly generating ambiguous situations.

In this paper, we develop a simple framework that overcomes the restrictive geometric, photometric and algorithmic constraints, mentioned before. We propose to study other kind of representations, based not only on local descriptors but also on global descriptors of the image, that we call *Intrinsic Images*, that simplify the (dense) matching process and can be used to generate new views from a stereo pair.

1.2 The Proposed Method

Our main goal consists of finding a representation with the following fundamental features:

1. The new representation can be represented through a simple image.
2. It must encode both the photometric and geometric structure of the original image.
3. The original images can be recovered again from this representation.
4. The disparity can be computed easily.
5. It can handle occlusion points.

To achieve that, we propose to use both local and global descriptors of the image in a new representation, so-called an *Intrinsic Image*. By using this representation, the computation of correspondences and disparity fields can be done in a straightforward manner, and all the requirements considered above are fulfilled.

However, some assumptions have to be made. Thus, our initial effort consists of defining three general and acceptable assumptions in order to produce a correspondence framework which can be considered sufficiently reliable.

- **Calibration** The stereo system is weakly calibrated, which means that corresponding epipolar lines are known.
- **Photometric Distortion** Given two corresponding points x_0 and y_0, the respective photometric values are related by a generic non-linear function $f(y_0) = \Psi(g(x_0))$, where $f(y_0)$ and $g(x_0)$ represent a given photometric measure (e.g. the brightness value). We study the cases where the photometric distortion function, $\Psi(g)$, is the identity (constant brightness) or affine (with contrast and mean brightness differences).
- **Geometric (perspective) Distortion** Two corresponding profiles are related by a disparity mapping function defined by $x = \Phi(y)$, between the two images. We assume that $\Phi(y)$ is a monotonic strictly increasing and differentiable function.

In summary, we propose to study the intensity-based matching problem of two corresponding epipolar lines, where the matched brightness points are ruled by a generic model for both geometric (perspective) and photometric distortion.

Additionally, there is a set of issues that we have to remark. First of all, the disparity mapping function $\Phi(y)$ includes important perspective and structure distortions of the scene. However the related assumptions made before are uniquely piecewise valid, considering that there are discontinuities and occlusion points on the images. Secondly, occlusion points have to be included coherently in our framework, by observing their position relatively to the corresponding points. However, unless prior knowledge exists, no depth information of an occlusion point can be recovered from two images. Finally, imposing $\Phi(y)$ to be strictly increasing, represents an order criteria, excluding order exchanges between two corresponding points. In fact an order exchange implies that an occlusion will occur between the two views. Thus, we consider those points as occlusions.

1.3 Structure of the paper

In Section 2, we consider the case of brightness constancy between the two views, without including occlusions, and define the *Intrinsic Images* and their most important properties. In Section 3, we introduce occlusions performing the necessary transformations to the proposed method. Finally, we generalize the method for the brightness distortion case. In the last two sections, we report some results on disparity estimation and generation of novel views, and discuss the general framework.

2 Intrinsic Images

The simplest kind of matching is that of two corresponding epipolar lines derived from two views of the same scene without occlusions. To simplify the general framework, we assume that there are no intensity changes due to viewing direction and that the disparity mapping function, $\Phi(y)$, defined as:

$$x = \Phi(y), \qquad \frac{d\Phi(y)}{dy} > 0 \tag{1}$$

verifies the order constraint and represents the unknown deformation at y to produce the corresponding point x. Assuming the brightness constancy hypothesis, the following nonlinear model can be expressed as

$$f(y) = g(x) = g(\Phi(y)) \tag{2}$$

In order to develop a simple matching framework, we propose alternative representations, based not only on local descriptors but also on global image descriptors.

The simplest example of a global descriptor of a scanline is the integral of brightness. One could associate each scanline pixel, x, to the sum of all brightness values between $x = 0$ and x. However, this integral would be different for two corresponding points in the two images, due to geometric (perspective) distortion.

In the remaining of this section, we will derive a new representation, based on a different global image descriptor. Using two horizontal scanlines of the perspective projection of the synthetic scene illustrated in Figure 1, we show that it can deal with perspective distortion between the two images.

Let f and g be the intensity values along corresponding epipolar lines. We now assume that both functions are differentiable, obtaining the following expression from the equation (2):

$$\frac{df(y)}{dy} = \frac{d\Phi(y)}{dy} \left.\frac{dg(x)}{dx}\right|_{x=\Phi(y)} \tag{3}$$

In the absence of occlusions, we can further assume that all brightness information is preserved between two corresponding segments $]y_1 \; y_2[$ and $]x_1 \; x_2[$ contained respectively in left and right images. Then, by assuming that $d\Phi(y)/dy > 0$ (order constraint), one proves analytically that the following equality holds:

$$\int_{y_1}^{y_2} \left|\frac{df(y)}{dy}\right| dy = \int_{y_1}^{y_2} \frac{d\Phi(y)}{dy} \left|\frac{dg(x)}{dx}\right|_{x=\Phi(y)} dy = \int_{x_1}^{x_2} \left|\frac{dg(x)}{dx}\right| dx \tag{4}$$

Now, let us define two functions α and β:

$$\alpha(y_i) = \int_{y_1}^{y_i} \left|\frac{df(y)}{dy}\right| dy \qquad \beta(x_i) = \int_{x_1}^{x_i} \left|\frac{dg(x)}{dx}\right| dx \tag{5}$$

where $x_i > x_1, y_i > y_1$ and $x_1,\ y_1$ are corresponding points. Equation (4) shows that when y_i and x_i are corresponding points, then $\alpha(y_i) = \beta(x_i)$. This means that, associating each scanline pixel, x, to the sum of the absolute value of the derivatives between $x = 0$ and x, we obtain the same values for corresponding points, independently of arbitrary image distortion, such as perspective effects.

In the following sections we will use these functions to build photometric and geometric image descriptors of a stereo pair. Such combined representation, called an *Intrinsic image*, will later be used for disparity estimation and synthesis of new views.

2.1 Photometric Descriptors

Using the stereo pair represented in Figure 1 as example, we can compute the functions $m_l = \alpha(y)$ and $m_r = \beta(x)$. We have shown in the previous section that

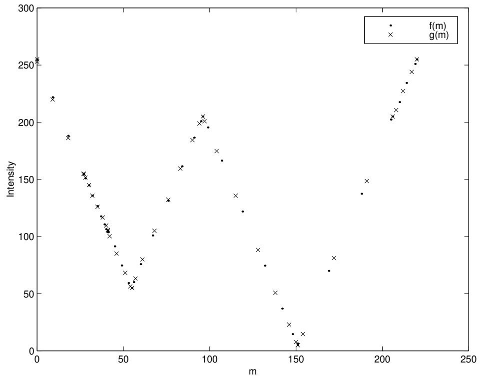


Fig. 2. Values from functions $m = \alpha(y)$, $m = \beta(x)$ versus the image intensities.

$m_l = m_r$, for two corresponding points. Going a step further, Figure 2 shows the values of these functions, $m_l = \alpha(y),\ m_r = \beta(x)$, versus the image intensity values, computed for the stereo images considered in the example. Not only the points are coincident, but also they can be represented through an image (putting together the various epipolar lines). One observes then a set of useful properties:

a) α and β are both monotonic increasing functions.
b) If $x_i,\ y_i$ are corresponding points, then $\alpha(y_i) = \beta(x_i)$, according to equations (4,5).
c) If $\alpha(y_i) = \beta(x_i)$, then $f(y_i) = g(x_i)$.
d) Let $m_l = \alpha(y)$. Every value of $m_l > 0$ corresponds to one and only one brightness value $f(y)$, meaning that the function $f(m_l)$ can be defined and represents a photometric descriptor. The same is applicable to $g(x)$.

These photometric descriptors, built from the stereo images, code photometric information available in the two images of the stereo pair, irrespective to perspective distortions. Hence in the absence of occlusions and under brightness constancy they are equal for the two images. Later, in Section 3, we will show how to use this property for occlusion detection.

When building a photometric descriptor for the image pair, we have lost information about the spatial domain that could lead to the computation of disparity. This aspect will be addressed in the following subsection.

2.2 Geometric Descriptors

We have seen a representation that codes all the photometric information of the stereo image pair, $f(m)$ and $g(m)$, and we now need to retrieve the geometrical information that is related to disparity.

Let us define the generalized functions $y'(m) = dy/dm$ and $x'(m) = dx/dm$. These functions, $x'(m)$ and $y'(m)$, are computed from images and take into account the local geometric evolution of the brightness along the scanlines.

Hence, we can form an image $x'(m)$ and $y'(m)$ by putting together various epipolar lines of an image. These descriptors convey all the necessary geometric (disparity) information available in the image pair. Each disparity, $d(x_i)$, and pixel value, x_i, can be recovered for each value of m_i, as follows:

$$(x_i, d(x_i)) = \left(\int_0^{m_i} x'(m)dm, \int_0^{m_i} (y'(m) - x'(m))dm \right) \tag{6}$$

We can generalize the definitions above for images with brightness discontinuities. In order to guarantee the validity of same theory, we define $df(y)/dy$ as a generalized function, admitting Dirac deltas in some points (at the brightness discontinuities). Thus, α is also discontinuous, $f(m)$ is uniquely defined for a restricted domain, and $x'(m)$ is zero in non-imaging areas (it means, values of m not represented by a brightness value).

The geometric descriptors, together with the photometric descriptors form complete *Intrinsic Images* that contain both photometric and geometric information, represented in the same coordinate system, of the stereo pair.

2.3 Definition of Intrinsic Images

In this section we will define formally the *Intrinsic Images* obtained from the photometric and geometric descriptors presented in the previous sections, and introduce some of the interesting applications of these representations.

Let k index corresponding epipolar lines of an stereo image pair. Then, the *Intrinsic Images*, $\mathcal{X}(m,k)$ and $\mathcal{Y}(m,k)$ are defined as:

$$\mathcal{Y}(m,\ k) = (f(m,k),\ y'(m,k)) \qquad \mathcal{X}(m,\ k) = (g(m,k),\ x'(m,k)) \tag{7}$$

where m is computed by equation (5) and $(m,\ k)$ are the coordinates of the intrinsic images. It is possible to reconstruct completely the original left and right images based on $\mathcal{Y}(m,k)$, $\mathcal{X}(m,k)$, respectively. Figure 3 shows the photometric and geometric components of the *Intrinsic Images*, computed for the stereo image pair shown in Figure 1.

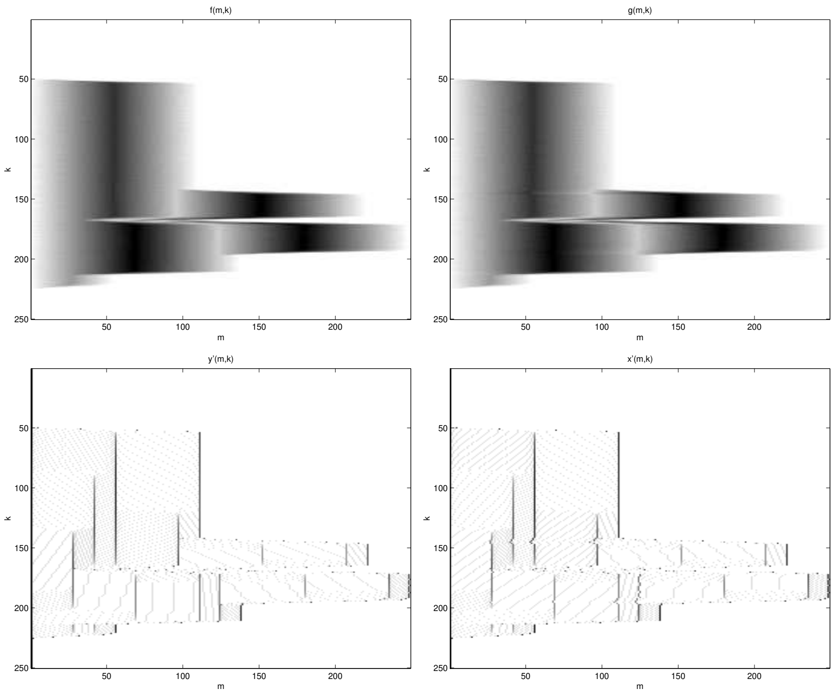


Fig. 3. Intrinsic images. Top: photometric descriptors. Bottom: geometric descriptors.

Given the properties described before, we state the following observation for Intrinsic Images, under the brightness constancy assumption and in the absence of occlusions:

Observation 1 — Intrinsic Images Property
If $f(y,k) = g(\Phi(y),k)$ and $d\Phi(y)/dy > 0$ for all $(y,\ k)$, then $f(m,k) = g(m,k)$ and the relation between $x'(m,k)$ and $y'(m,k)$ gives the geometric deformation between corresponding points.

From this observation we derive an interesting property, that will allow us to generate novel views of the observed scene, including all perspective effects.

Observation 2 — Synthesis of new Images
Assume that both cameras are parallel, related by a pure horizontal translation T, with intrinsic images $\mathcal{Y} = (f(m,k),\ y'(m,k))$ and $\mathcal{X} = (f(m,k),\ x'(m,k))$. Then, views at intermediate positions jT (where $0 \leq j \leq 1$) have the following

intrinsic images, $\mathcal{I}_j$*:*

$$\mathcal{I}_j(m,\ k) = (f(m,k),\ jx'(m,k) + (1-j)y'(m,k)) \tag{8}$$

Proof. Suppose that the disparity between two corresponding points x_j and y of two generic parallel cameras is given by the well known expression $x_j = y + jT/Z$, where Z denotes the depth of the object relatively to the cameras and jT represents the translation between them. Derivating both terms of the equality in relation to m, we obtain

$$x_j^{'}(m) = y'(m) + jT\frac{dZ^{-1}}{dm} \tag{9}$$

Finally, by performing a weighted average between the left and the right cameras ($j = 0$ and $j = 1$ respectively), we obtain the same expression for $x_j^{'}(m)$:

$$x_j^{'}(m) = jx'(m) + (1-j)y'(m) = y'(m) + jT\frac{dZ^{-1}}{dm} \tag{10}$$

$\square$

This result provides a means of generating intermediate unobserved views simply by averaging the geometric component of the original *Intrinsic Images*. It accounts for perspective effects without an explicit computation of disparity.

We have described the essential framework for our matching approach assuming brightness constancy and in the absence of occlusions. We have defined *Intrinsic Images* and shown how to use these images to compute disparity in a direct way. Next, we will introduce occlusion information and, finally, relax the brightness constancy constraint.

3 Dealing with Occlusions

In computer vision, occlusions are often considered as artifacts, undesirable in many applications, ignored in others. However, occlusions are neither rare nor meaningless in stereo interpretation. Anderson and Nakayama [1] have shown that an occlusion plays an important role for the human perception of depth and motion, and [7] describes a theoretical connection between occlusion classification and egomotion detection in computer vision.

Many algorithms have been designed to handle occlusions for multiple image motion or disparity estimation [5, 9, 3]. Here, we focus on introducing occlusions in *Intrinsic Images*, ensuring that there is no exceptional treatment based on cost functions or adhoc thresholds.

An occlusion occurs when a surface is in front of an occluded region, which can be seen by only one camera. However, unless we impose prior models to image disparities, 3D structure or to global image features, we can detect the existence of a local occlusion based on photometric information. We will show how to include occlusion information in *Intrinsic Images*, based on the theory developed in the last section.

First of all, even in the presence of occlusion points, an intrinsic image can be defined as before. However, the *Intrinsic Images Property* stated in Observation 1, is not verified because its sufficient condition is not verified. In fact, the condition $f(y,k) = g(\Phi(y),k)$ is only piecewise valid (along corresponding profiles), but it is not valid in general, namely in occlusion points.

It is worth noticing important differences between the usual cartesian images and the associated photometric descriptor of the intrinsic images, in a stereo pair. While the intrinsic images will only differ in the presence of occlusions, cartesian images differ both by perspective effects and occlusions. This means that, in order to detect occlusions or, equivalently, photometric dissimilarities, we can rely on the photometric descriptors of the intrinsic images, where the geometric distortions have been removed.

Therefore, we propose to define global image descriptors similar to those discussed previously. Consider

$$m_l = \alpha(y_i) = \int_{y_1}^{y_i} \left| \frac{df(y)}{dy} \right| dy \qquad m_r = \beta(x_i) = \int_{x_1}^{x_i} \left| \frac{dg(x)}{dx} \right| dx \quad (11)$$

computed on corresponding epipolar lines of the left and right cameras, where x_1 and y_1 are the respective initial (and not necessarily corresponding) points.

In the previous section, we have shown that in the absence of occlusions, $m_l = m_r$ for corresponding points, greatly simplifying the matching procedure. In the presence of occlusions, the matching is not that trivial, but m_l and m_r can still be related by a simple function. Let m_l and m_r be parameterized by t as follows:

$$m_l = r(t) \qquad m_r = s(t) \quad (12)$$

The curve produced by these functions can yield uniquely three forms:

1. Horizontal inclination ($dr(t)/dt = 1$; $ds(t)/dt = 0$): there exists a region in the left camera, that is occluded in the right camera.
2. Vertical inclination ($dr(t)/dt = 0$; $ds(t)/dt = 1$): there exists a region in the right camera, that is occluded in the left camera.
3. Unitary inclination ($dr(t)/dt = 1$; $ds(t)/dt = 1$): both profiles match (no occlusions).

Figure 4 shows examples of two (m_l, m_r) matching scenarios with and without occlusions.

Hence, the problem to solve is that of determining the mapping curve from m_l to m_r, which can be done through several approaches. Fortunately, there is a low cost algorithm that solves it optimally, in the discrete domain.

Assume that we have two sequences $F = \{f(m_{l1}),\ f(m_{l2}), ...,\ f(m_{lp})\}$ and $G = \{g(m_{r1}),\ g(m_{r2}), ...,\ g(m_{rq})\}$ given by the photometric information along corresponding epipolar lines of the left and right intrinsic images. The corresponding points constitute a common subsequence of both F and G. Moreover, finding the set of all corresponding points corresponds to finding the maximum-length common subsequence of F and G. This problem consists in the well known

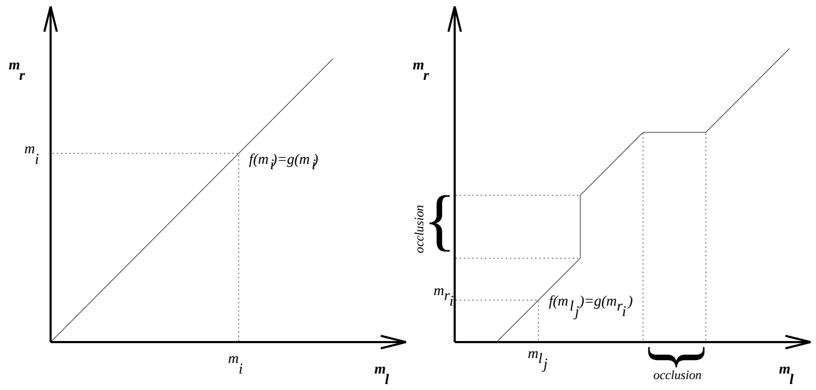


Fig. 4. Matching scenarios without (left) and with occlusions (right).

longest-common-subsequence (LCS) problem [2], which can be solved efficiently using dynamic programming.

Notice that an LCS solution for our problem implies two things: (1) the corresponding points obey to the order constraint; (2) an order exchange represents existence of occlusions. These two implications can produce some ambiguity situations. However, such cases correspond mostly to perceptual ambiguities only solved with prior knowledge about the scene.

After finding an LCS solution, or, equivalently, the curve that matches m_l and m_r, we can change the definition of the *Intrinsic Images* in order to maintain the theory described in last section.

Observation 3 — Intrinsic Images Property with occlusions
Given two stereo images, if $f(y,k) = g(\Phi(y),k)$ and $d\Phi(y)/dy > 0$ for almost all $(y,\ k)$ (except for occlusions points), then it is possible to determine a pair of intrinsic images $\tilde{\mathcal{Y}}(t,k)$ and $\tilde{\mathcal{X}}(t,k)$.

$$\tilde{\mathcal{Y}}(t,k) = (\tilde{f}(t,k),\ \tilde{y}'(t,k)) \qquad \tilde{\mathcal{X}}(t,k) = (\tilde{g}(t,k),\ \tilde{x}'(t,k)) \tag{13}$$

where $\tilde{f}(t,k) = \tilde{g}(t,k)$ and the relation between $\tilde{x}'(t,k)$ and $\tilde{y}'(t,k)$ gives the geometric deformation between corresponding and occlusion points of the stereo images. Knowing the functions $r(t)$ and $s(t)$ of equation (12), these new intrinsic images are found based on the original intrinsic images (as presented in Section 2), by performing the following transformation:

$$(\tilde{f}(t,\ k),\ \tilde{y}'(t,k)) = \begin{cases} (f(r(t),\ k),\ y'(r(t),k)) & if \quad dr(t)/dt = 1 \\ (g(s(t),\ k),\ 0) & if \quad dr(t)/dt = 0 \end{cases}$$
$$(\tilde{g}(t,\ k),\ \tilde{x}'(t,k)) = \begin{cases} (g(s(t),\ k),\ x'(s(t),k)), & if \quad ds(t)/dt = 1 \\ (f(r(t),\ k),\ 0) & if \quad ds(t)/dt = 0 \end{cases} \tag{14}$$

This observation has two important implications. First, by computing the functions r and s as a solution of the LCS problem, we can derive a coherent framework which permits to compute disparities or generate new views as defined in last section. Secondly, the generated intermediate views exhibit consistent information in occlusion areas. However, it does not imply that this information is consistent with the real 3D structure of the occluded points (given the impossibility to recover that structure).

4 Photometric Distortion

So far we have considered the hypothesis of brightness constancy. In a real stereo system, however, the brightness can change due to viewing direction or different specifications of the cameras. A convenient model to account for photometric distortion is the following:

$$f(y) = a \cdot g(\Phi(y)) + b, \qquad a > 0 \tag{15}$$

This model represents an affine transformation, where a and b are the difference in contrast and brightness between the two images. Some authors prefer to estimate a priori the affine parameters [4], before applying the correspondence procedure. This can be performed by analyzing the global intensity function and its derivatives. However, it would be constructive to study the geometric influence of the affine distortion on the intrinsic image structure.

Considering the brightness distortion in Equation (15), the effect of bias, b, can be eliminated by preprocessing both signals with a zero-mean filter. Thus, we assume that the contrast different term a is the dominant term.

By assuming that $f(y) = a \cdot g(\Phi(y))$ and applying equations (4), a simple relation between the horizontal axis of the intrinsic images is found: $m_l = a \cdot m_r$. This means that the photometric information along the scanlines of the left intrinsic image is scaled along the horizontal axis and in amplitude (the intensity values), by a, with respect to the right intrinsic image. Thus, the geometric deformation induced by the brightness distortion in the intrinsic images, is ruled by the following equality:

$$\frac{f(m_l)}{m_l} = \frac{g(m_r)}{m_r} \tag{16}$$

We could apply directly to the intrinsic images a correspondence procedure by using the equation (16). Nevertheless this implies some search effort.

Another solution to overcome this problem consist of transforming the brightness function by a simple function which exhibits some invariant properties related to linear distortions. The logarithm function is a good candidate. In fact, when y and x are corresponding points, we have:

$$\frac{d \log |f(y)|}{dy} = \frac{d\Phi(y)}{dy} \frac{d \log |g(x)|}{dx} \tag{17}$$

defined wherever $f(y)$ and $g(x)$ are different of zero. Applying this relation to equation (4), one can conclude that it is still possible to define coherent photometric and geometric descriptors, since $m_l = m_r$ (in absence of occlusions). This means that, in general, the intrinsic image theory remains applicable.

5 Results

Along the paper, we have used a simple synthetic stereo pair in order to illustrate the various steps of our approach. By using the intrinsic images shown in Figure 3, we have used Equation (6) to compute a dense disparity map, shown in Figure 5, without requiring any search algorithms. We applied the same methodology to another stereo pair, from the baseball sequence, and computed the respective disparity map, shown in Figure 5.

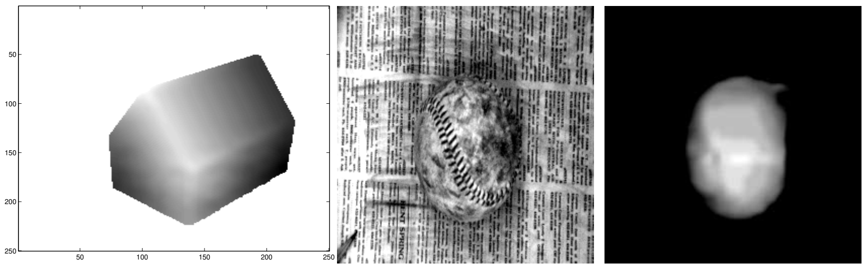


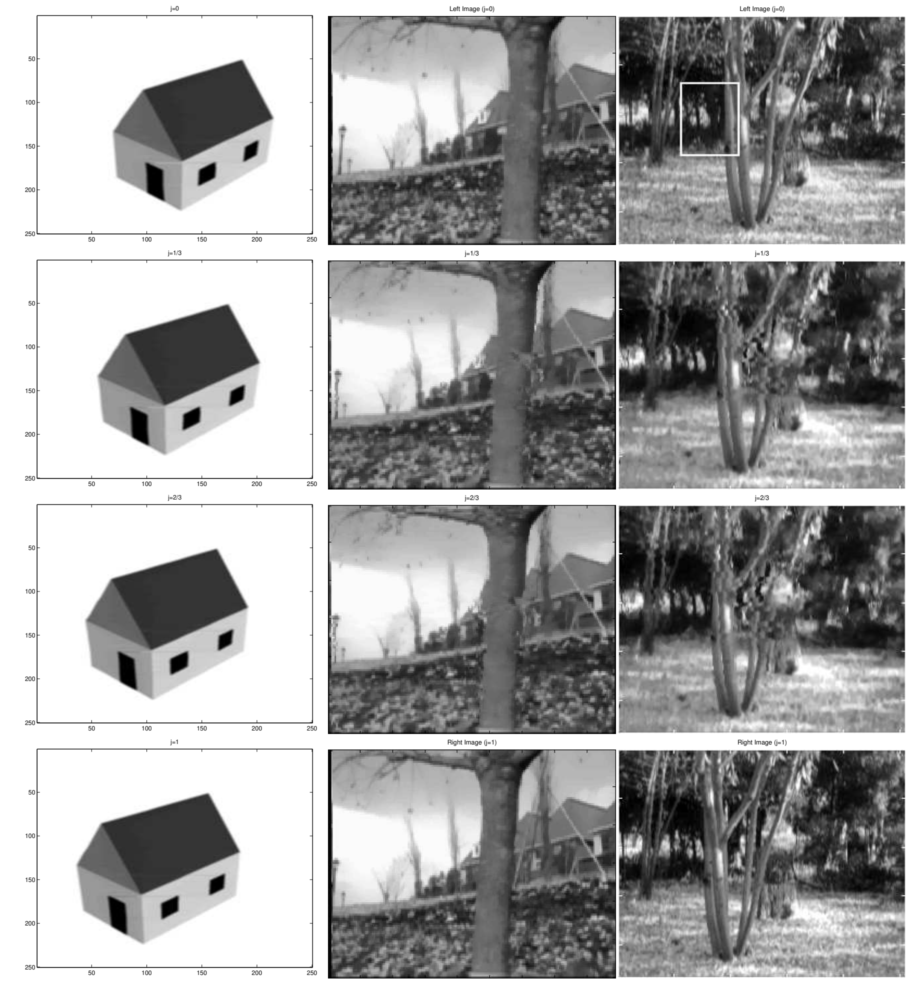
Fig. 5. Disparity map determined directly from the Intrinsic Images. On the left: Disparity map of the synthetic pair; On the center: Left image of a stereo pair from the baseball sequence; On the right: the associated disparity map.

In Figures 6a, we show results obtained with the Synthetic Images, "flower garden" and "trees" sequences. Except in the synthetic case, a significant amount of occlusions are present. We have applied the proposed method, based on the *Intrinsic Images* to synthesize new views from the original left and right stereo images. The occluded regions move coherently in the sequences. The perceptual quality of the reconstructed images is comparable with that of the original ones. Notice that in the "trees" sequence there is an area, roughly on the center of the image, where some occlusions associated to order exchanges create erroneous solutions for the synthesis of that area. This can only be solved introducing more images of the sequence or with prior knowledge about the scene. To illustrate the ability of the method to cope with occlusions, Figure 6b shows the recovered views on a detail of the images (see little square in the respective sequence), with a strong discontinuity in the depth. As expected, the occluded points disappear coherently behind the tree, without any smoothing.

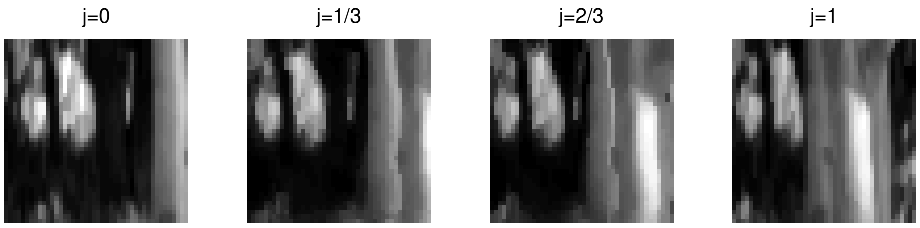
These examples illustrate how *Intrinsic Images* deal with perspective effects and occlusions, generating dense disparity fields and novel views from an original stereo pair.

6 Conclusions

We have proposed a new image representation - *Intrinsic Image* which allows for solving the intensity-based matching problem under a realistic set of assump-



a)



b)

Fig. 6. a) Synthesis of intermediate views for the synthetic images (left), the Flower Garden Sequence (center) and for the Tree Sequence (right). b) The evolution of a detail of the Tree Sequence (see little square in the respective sequence) with occlusions.

tions. The concept of intrinsic images is a useful way to approach the problem of stereo vision, and leads to a straightforward computation of disparity and new views.

Intrinsic Images of a stereo pair give exactly the complete photometric and geometric structure of the 3D scene, independently of the observed geometric (perspective) distortion. Secondly, the occlusions are introduced naturally in this framework, which means that we apply a consistent interpretation to the occluded points.

An Intrinsic Image is composed by a photometric and a geometric descriptor, which contain all the necessary information to reconstruct directly the original images, disparity values and other views of the same scene. We have presented some results with real stereo images. The method is very robust to the existence of occlusions and reveals high performance in multi-view synthesis.

This approach is a powerful tool to cope not only with a stereo pair of images but also with a sequence of images, both in the discrete and continuous sense. In the future, we plan to study the potentialities of this approach applied to a larger number of images, namely in dense optical flow computation and egomotion estimation. We plan also to develop an automatic occlusion characterization based on a large number of images of same scene, in order to facilitate the creation of the associated intrinsic images with occlusion information.

Acknowledgments

This work was partially funded by the EU ESPRIT-LTR proj. 30185.NARVAL and a grant from the PRAXIS-XXI research program.

References

1. B. L. Anderson and K. Nakayama. Toward a general theory of stereopsis: Binocular matching, occluding contours, and fusion. *Psych. Review*, 101:414–445, 1994.
2. T. Cormen, C. Leiserson, and R. Rivest. *Introduction to Algorithms.* The MIT Press, 1990.
3. U. R. Dhond and J. K. Aggarwal. Stereo matching in the presence of narrow occluding objects using dynamics disparity search. *PAMI*, 17(7):719–724, July 1995.
4. R. Haralick and L. Shapiro. *Computer and Robot Vision*, volume 2. Addison-Wesley, 1993.
5. Y. Ohta and T. Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *PAMI*, 7(2):139–154, 1985.
6. C. Silva and J. Santos-Victor. Robust egomotion estimation from the normal flow using search subspaces. *PAMI*, 19(9):1026–1034, September 1997.
7. C. Silva and J. Santos-Victor. Motion from occlusions. In *Proc. of 7th International Symposium on Intelligent Robotic Systems*, 1999.
8. C. Tomasi and R. Manduchi. Stereo without search. In *Proc. of European Conference on Computer Vision*. Springer-Verlag, 1996.
9. J. Weng, N. Ahuja, and T. Huang. Matching two perspective views. *PAMI*, 14(8):806–825, 1992.