

Research Article

An Occlusion Approach with Consistency Constraint for Multiscopic Depth Extraction

Cédric Niquin,^{1,2} Stéphanie Prévost,¹ and Yannick Remion^{1,2}

¹CRESTIC SIC, Université de Reims Champagne-Ardenne, 51100 Reims, France

²3DTV Solutions, 8 rue Gabriel Voisin, 51100 Reims, France

Correspondence should be addressed to Cédric Niquin, cedric.niquin@univ-reims.fr

Received 30 April 2009; Revised 31 August 2009; Accepted 28 November 2009

Academic Editor: Nikolaos Grammalidis

Copyright © 2010 Cédric Niquin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This is a new approach to handle occlusions in stereovision algorithms in the multiview context using images destined for autostereoscopic displays. It takes advantage of information from all views and ensures the consistency of their disparity maps. We demonstrate its application in a correlation-based method and a graphcuts-based method. The latter uses a new energy, which merges both dissimilarities and occlusions evaluations. We discuss the results on real and virtual images.

1. Introduction

Augmented reality has many applications in several domains such as games or medical training. On the other hand autostereoscopic display is an emergent technology, which adds a perception of depth enhancing the users immersion. Augmented reality can be applied to autostereoscopic display in a straightforward way by adding virtual objects on each image. However, it is much more interesting to use the depth-related information of the real scene so that virtual objects could be hidden by real ones.

To that end, we need to obtain one depth map for each view. The particular context of images destined to be viewed on autostereoscopic displays allows us to work on a simplified geometry (e.g., no rectification is needed, epipolar pairs are horizontal lines of the same rank, and disparity vectors are thus aligned along the abscissa). However, our aim is to obtain good assessment of depth in all kinds of scenes, without making any assumption on their contents. Indeed images may have homogeneous colors as well as they may have various colors. Also, due to the principle of autostereoscopic displays, the users can see two images at the same time. It is then crucial to have strongly consistent depth maps. For example if a virtual object is drawn in front of a real object in one view, it has to be drawn in the same order in all views. Therefore we introduce a new occlusion approach

for multiview stereovision algorithms, which aims to ensure the consistency of the depth maps.

We propose an example of application of our approach in a correlation-based method and in a symmetrical graphcuts-based method. Finally we discuss their results.

2. Related Work

Stereovision algorithms aim to find the disparity maps in order to deduce the depth maps. That is the reason why we will use the phrase “disparity maps” instead of “depth maps” in the following lines. Depth maps can be easily obtained from disparity maps using a triangulation step.

Let us admit we have a set of N images, numbered from the left (0) to the right ($N - 1$), shot with a parallel capture system specifically designed for autostereoscopic displays. Figure 1 illustrates these shooting conditions with $N = 4$. f is the focal distance and d_{i0c} is the Distance Intra Optical Centers. I_i designates the set of pixels of the i th image. $d_i(p)$ is a function associating a disparity vector with any pixel p from I_i . This vector is the difference between the coordinates of the corresponding pixel of p in image $i + 1$ and those of p in image i . The corresponding pixel of p in the next image is then at position $p + d_i(p)$. Since we are using a simplified geometry, $d_i(p)$ is of dimension one. It is an integer too

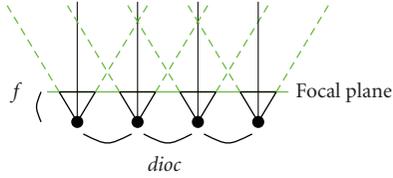


FIGURE 1: Illustration of the parallel camera configuration with four optics (illustrated by black dots).

($d_i(p) \in \mathbb{Z}$), so that we do not have to deal with subpixels. Moreover it is possible to find the corresponding pixel of p in any image using the only disparity $d_i(p)$. In any k image, the corresponding pixel is given by $p + (k - i)d_i(p)$. For example the corresponding pixel of p in the previous image is $p - d_i(p)$.

Optical flow algorithms are based on a cost $L(p, q)$, which evaluates the color dissimilarity between two pixels p and q . In the case of color images, it is given by

$$L(p, q) = |r_p - r_q| + |g_p - g_q| + |b_p - b_q|, \quad (1)$$

where r_p, g_p , and b_p (resp., r_q, g_q , and b_q) are the red, green, and blue components of p (resp., q). Several methods use the multiview aspect in order to make such algorithms more robust and less sensitive to noise. The local cost M_α^i of a pixel $p \in I_i$ according to disparity α is then given by

$$M_\alpha^i(p) = \sum_{k=0}^{N-2} L(p + (k - i)\alpha, p + (k + 1 - i)\alpha). \quad (2)$$

The reader can refer to Scharstein and Szeliski [1] for a complete taxonomy of dense stereovision algorithms. In many recent publications from this domain, authors use color segmentation in their methods [2–5]. However color segmentation and other primitive extraction methods are independent for each view. It is then impossible to ensure the consistency between the disparity maps. Moreover, it may be a cause of errors when applied to images with mainly homogeneous colors. Therefore these methods are incompatible with the objective presented previously, which imposes working in a local colorimetric context. We may not have any assumption about the content of the images and, thus, we cannot extract other features. So we aim to make up for this lack of information by taking advantage of redundancies in the N images.

A lot of algorithms deal with occlusions in order to obtain better disparity maps, which preserve discontinuity at object's boundaries. The first step to deal with occlusions is to be able to detect them. Egnal and Wildes [6] compare five approaches. Some of them are based on the idea that color discontinuities correspond to depth discontinuities. These approaches are called photometry-based approaches. Alvarez et al. [7] use the gradient of the gray levels in order to locally adjust the smoothness constraint of their energy: the lower the value of the gradient, the stronger the smoothness constraint. On the other hand, geometry-based approaches use disparities in order to detect occlusion

areas. The reader can refer to [6] for a complete comparison of such methods. We prefer geometry-based approaches to the photometry-based ones since they do not make any assumption on colors and allow disparity maps from the N images to interact and to be linked. This link ensures the consistency of the disparities and makes up for the lack of information previously discussed. The most widely used geometry-based method is the Left Right Checking (LRC) approach [8–10]. The principle is that a pixel should match a pixel from another image with the same disparity; otherwise an occlusion occurs. In the case of two images (numbered 0 and 1), this can be expressed by

$$\begin{aligned} \epsilon_0(p) &= ||d_0(p) - d_1(p + d_0(p))||, \\ \epsilon_1(p) &= ||d_1(p) - d_0(p - d_1(p))||, \end{aligned} \quad (3)$$

where $\epsilon_i(p)$ is close to zero when there is no occlusion and high when pixel p is occluded in image i . There have been several attempts to improve the robustness of this method [11, 12]. However, the original LRC is still the most popular approach [2, 5, 13, 14]. We propose an LRC-based approach, which differs in several points:

- (i) it is extended to the multiview context,
- (ii) it ensures a geometric consistency between depth maps.

After the detection step, the main difficulty is to handle occlusions in the matching algorithms. Woetzel and Koch [15] propose a correlation-based algorithm, which do not add up dissimilarity costs for the whole set of image pairs but for a subset of it. They replace (2) by

$$M_\alpha^i(p) = \sum_{(j,k) \in S} L(p + (j - i)\alpha, p + (k - i)\alpha), \quad (4)$$

where S is the set of chosen image pairs. The authors propose two methods to choose the set S of image pairs. The first one is to select the furthest left or the furthest right images. The second one is to select the pairs with the smallest L costs. This method reduces the impact of occlusions on the results but introduces a lot of errors in the images which are the furthest away.

There are two categories of methods based on energy minimization performing the matching while taking occlusions into account.

The first category contains iterative methods [8, 10] based on Algorithm 1. In order to apply this principle, Proesmans et al. [8], who work with two images, use four maps, one disparity map plus one occlusion map per image. The occlusion maps are computed using the LRC approach. Strecha and Van Gool [10] extend this principle to the multiview context. First, N disparity maps are computed from the N views. Then for each view, $N - 1$ occlusion maps are computed as being the LRC evaluation with all other views.

In order to obtain better results, some methods start again at step 2 when step 3 is over and loop until the system converges. The problem of iterative methods is that

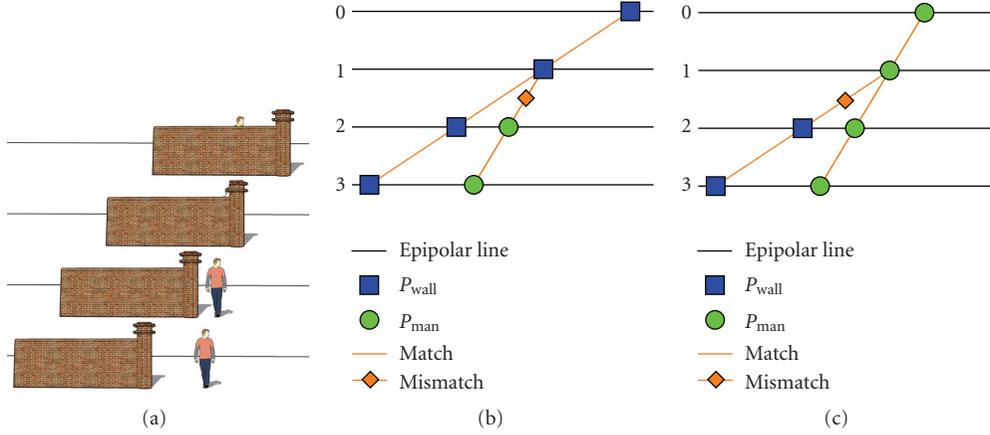


FIGURE 2: Examples of matching graphs.

- (1) disparities estimation without taking into account occlusions
 - (2) estimation of occlusion areas
 - (3) disparities estimation with occlusion handling

ALGORITHM 1: Iterative methods principle.

disparities and occlusions estimations are independent, do not interact with each other and, thus, do not ensure a global geometric consistency.

The second category is then composed of methods to estimate occlusions and disparities simultaneously. In the context of two views, Alvarez et al. [9] introduce the following energy $E(d_0, d_1) = E_0(d_0, d_1) + E_1(d_0, d_1)$ with

$$\begin{aligned} E_0(d_0, d_1) &= E_0^d(d_0) + \chi E_0^s(d_0) + \beta E_0^o(d_0, d_1), \\ E_1(d_0, d_1) &= E_1^d(d_1) + \chi E_1^s(d_1) + \beta E_1^o(d_0, d_1), \end{aligned} \quad (5)$$

where E_k^d evaluates dissimilarities between corresponding pixels, E_k^s corresponds to the smoothness constraint, and E_k^o is the sum of ϵ_k for every pixel of image k . χ and β are weighting factors.

Note that even if pixels are detected as occluded, their dissimilarities are still taken into account in the dissimilarity term. That means that this term contains dissimilarities of mismatching pixels, which have nothing in common. This is a problem since that introduces noise into the energy. In order to solve that, Ince and Konrad [13] use an energy similar to (5) with the dissimilarity terms of the form

$$\begin{aligned} E_0^d(d_0) &= \sum_{p \in I_0} F(\epsilon_0(p)) \times L(p, p + d_0(p)), \\ E_1^d(d_1) &= \sum_{p \in I_1} F(\epsilon_1(p)) \times L(p, p - d_1(p)), \end{aligned} \quad (6)$$

where F is a weighting function, which approaches zero when ϵ is high (i.e., occlusions are detected). Moreover, they use

the smoothness term in order to extrapolate the disparities in occluded areas.

By the same token, we have proposed a multiview graph-cuts-based method in [14], which integrates occlusion penalties in its energy function without integrating dissimilarities of mismatching pixels.

In spite of the fact that these methods use smooth and discontinuity preserving functions, they still can contain inconsistencies that we will detail in Section 3.

3. A New Approach for Occlusion Detection

3.1. Overview. Let us imagine a standard scene with a man behind a wall. Four views of this scene are shot; Figure 2(a) shows corresponding epipolar lines from each image and superimposes them. This representation, that we call matching graph, is useful in order to define matches and occlusions. In Figure 2(b), pixel $p_{\text{man}} \in I_3$ (green circle) has disparity $d_3(p_{\text{man}})$. The corresponding pixel in I_2 is also a pixel of the man, so there is a match between these pixels. In image 1, the corresponding pixel of p_{man} is part of the wall (blue square). It has a larger disparity (since the wall is nearer). Therefore there is an occlusion between these two pixels, as shown in Figure 2(b) with an orange diamond. The scene described in Figure 2(c) is an example of an impossible matching graph. This graph shows that the wall pixel in image 2 corresponds to a pixel in image 1, which is not a pixel of the wall but a representation of the man. It has a smaller disparity. This situation is impossible as it supposes that the wall is hidden by the man whose disparity implies that he is behind it. This is what we call a consistency error. The LRC does not handle this kind of inconsistencies.

We propose the following rules in order to define our approach of occlusions. Let us assume $p \in I_i$ and its corresponding pixel $q \in I_j$:

- (i) if $d_i(p) = d_j(q)$, p and q match,
- (ii) if $d_i(p) < d_j(q)$, p is occluded,
- (iii) if $d_i(p) > d_j(q)$, there is a consistency error.

In order to simplify writing, we call θ the occlusion image when an occlusion occurs at image number θ , that is, when corresponding pixels from images θ and $\theta + 1$ do not match. In Figure 2(b), for instance, there is an occlusion between images 1 and 2, θ is then equal to 1.

3.2. Energy Function. In order to take the rules presented above into account, we use an energy function of the form

$$E(d) = E^m(d) + \sum_{i=0}^{N-1} E_i^s(d_i), \quad (7)$$

where d is the set of all disparity functions d_i . E_i^s is the smoothness term, which may be the same as the one used in (5). This smoothness constraint is applied to each disparity function. E^m contains all the dissimilarity, occlusion, and consistency penalties.

In the case of Figure 2(b), E^m will include the three dissimilarities between the four pixels of the wall, plus one dissimilarity between the two pixels of the man, plus one occlusion penalty between the two mismatching pixels. In Figure 2(c), E^m is the sum of the three dissimilarities between the man pixels, the dissimilarity between the wall pixels, and the consistency penalty. Of course these examples take a very small number of pixels into account whereas E^m considers all of them.

Finally, this term is given by

$$E^m(d) = \sum_{i=0}^{N-2} \sum_{p \in I_i} E_{i+}^{\text{local}}(d, p, p + d_i(p)) + \sum_{i=1}^{N-1} \sum_{p \in I_i} E_{i-}^{\text{local}}(d, p, p - d_i(p)), \quad (8)$$

where E^{local} is the local cost between $p \in I_i$ and its corresponding pixel either in the previous ($i-$) or the next ($i+$) image. Let us call this pixel q . We have

$$E_{i+}^{\text{local}}(d, p, q) = \begin{cases} L(p, q) & \text{if } d_i(p) = d_{i+1}(q), \\ K^{\text{occ}} & \text{if } d_i(p) < d_{i+1}(q), \\ K^{\text{con}} & \text{if } d_i(p) > d_{i+1}(q). \end{cases} \quad (9)$$

The term $E_{i-}^{\text{local}}(d, p, q)$ is the same except that q is in image $i - 1$ instead of $i + 1$. K^{occ} and K^{con} are two constant values corresponding to the occlusion penalty and the consistency penalty, respectively. Due to our specific domain of application (augmented reality on autostereoscopic displays), the consistency constraint must be very strong and the value of K^{con} is then very high to ensure that this case will never happen.

This energy function can be used in different methods as we will see in Section 4.

4. Application

In this section, we present two applications of our occlusion approach. The first one does not use any smoothness constraint and focuses on our approach of occlusions in order to

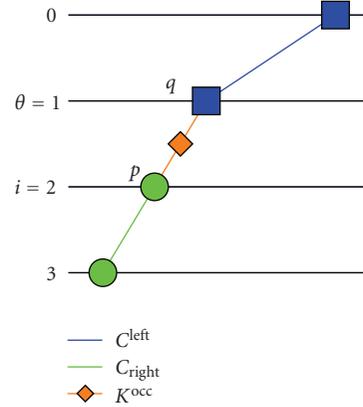


FIGURE 3: Example of the components of C^{local} (10).

emphasize its relevance on a correlation-based method. The second one is an application of the energy function as defined in the previous section on a graph-cuts-based method.

Both methods use the same constant K^{occ} . We found empirically that a value of 100 gives good results with our different sets. We give K^{con} a value of 3000.

4.1. Correlation-Based Method. This method uses two distinct local costs. The first one supposes there is no occlusion and the second one supposes there is exactly one occlusion. These two costs are in competition by means of a Winner Takes All (WTA) algorithm.

The first cost could be any local cost as found in the literature. Our implementation uses cost M_α^i described in (2) where L is the absolute difference of intensities summed over the three color components.

The second cost C^{local} is a local subset of E^m (8). Indeed the cost of a p pixel includes the E^{local} (9) energies for all pixels linked to p in the N images, assuming that there is only one occlusion and two disparities (on the left and the right of the occlusion). In order to ensure the consistency of the matching graph implicitly, only values meeting the consistency condition are tested. Therefore the constant K^{con} never appears. C^{local} finally contains one penalty K^{occ} due to the occlusion, plus dissimilarities between matching pixels. The local cost for p in image i is

$$C_i^{\text{local}}(p, \theta, \delta_{\theta-}, \delta_{\theta+}) = C_i^{\text{left}}(q, \theta, \delta_{\theta-}) + C_i^{\text{right}}(q, \theta, \delta_{\theta+}) + K^{\text{occ}}, \quad (10)$$

where θ is the occlusion image and $\delta_{\theta-}$ and $\delta_{\theta+}$ are two disparities, respectively, on the left and on the right of the occlusion. Figure 3 shows the three terms of C^{local} with $i = 2$ and $\theta = 1$ with the same configuration as in Figure 2(b), that is, it shows the costs of each term. q is the corresponding pixel of p in image θ : q is equal to $p + (\theta - i)\delta_{\theta-}$ if i is on the left of the occlusion ($i < \theta$), and equal to $p + (\theta - i)\delta_{\theta+}$ if it is on the right ($i > \theta$). C^{left} and C^{right} contain dissimilarity costs on

the left and on the right sides of the occlusion, respectively. They are given by

$$C_i^{\text{left}}(q, \theta, \delta_{\theta-}) = \sum_{k=0}^{\theta-1} L(q+(k-\theta)\delta_{\theta-}, q+(k+1-\theta)\delta_{\theta-}),$$

$$C_i^{\text{right}}(q, \theta, \delta_{\theta+}) = \sum_{k=\theta+1}^{N-2} L(q+(k-\theta)\delta_{\theta+}, q+(k+1-\theta)\delta_{\theta+}).$$
(11)

In order to ensure the consistency of the matching graph, only disparities meeting the following condition are tested:

$$\begin{aligned} \delta_{\theta-} &< \delta_{\theta+} & \text{if } i < \theta, \\ \delta_{\theta-} &> \delta_{\theta+} & \text{if } i > \theta. \end{aligned}$$
(12)

Finally, the selection is based on a WTA algorithm: if the minimum cost for a pixel is obtained using M_{α}^i , then the disparity α is assigned to the pixel. If it is obtained using C^{local} then the disparity assigned to the pixel is either $\delta_{\theta-}$ or $\delta_{\theta+}$, depending on whether i is, respectively, on the left or on the right of the occlusion.

4.2. Graph-Cuts-Based Method. Our method is based on the energy function previously described in (7) and (8). We use the graph-cuts method in order to minimize our energy. Please refer to publications by Boykov et al. [16] for a complete presentation of the graph-cuts method and by Kolmogorov and Zabih [17] for an explanation of the graph construction. We use the α -expansion algorithm and, unlike others, we loop only once for each disparity. Moreover we always loop from the highest disparity to the lowest one, since we have found that this gives more accurate occlusion areas in the results. The graph we present in this section is based on this assumption, as we will see further. Our graph is composed of one node for each pixel from all the images. It also has a source (s) and a sink (t), which mean “keep the same disparity” and “change to disparity α ”, respectively.

Now, we will see how to construct the graph corresponding to E^{local} (9). In fact p will not have the same corresponding pixel in the other image whether it changes its disparity or not. The graph corresponding to E^{local} is then composed of three nodes p , q_s , and q_t , which are the corresponding pixels if p is cut from the source or from the sink. However, it can be separated into two graphs using pairs (p, q_s) and (p, q_t) . Figure 4 shows the general graph corresponding to E^{local} .

(i) *Pair* (p, q_s)

This pair makes sense only if p is cut from the source. It means that cuts $C_{ts}^{pq_s}$ and $C_{tt}^{pq_s}$ are equal to 0. Cut $C_{ss}^{pq_s}$ is equal to either $L(p, q_s)$ or K^{occ} depending on the previous disparities of p and q_s . Since our main loop is going from the highest disparity to the lowest one, cut $C_{st}^{pq_s}$ means q_s will have a lower disparity than p . It is then equal to K^{con} . As mentioned by Kolmogorov and Zabih [17], the

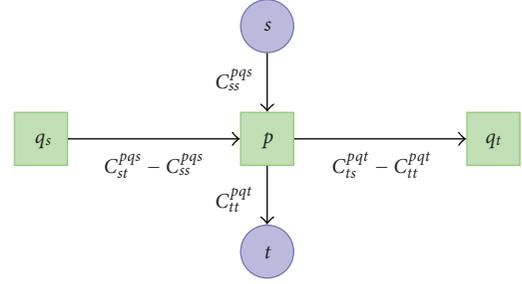


FIGURE 4: Graph corresponding to E^{local} (9).

following constraint has to be checked in order to ensure that the graph is achievable:

$$C_{st} + C_{ts} \geq C_{ss} + C_{tt}. \quad (13)$$

In our case $C_{ts}^{pq_s}$ and $C_{tt}^{pq_s}$ are equal to 0 and $C_{st}^{pq_s}$ is always greater than $C_{ss}^{pq_s}$, so that our graph is achievable.

(ii) *Pair* (p, q_t)

On the same principle as the previous pair, we have $C_{ss}^{pq_t}$ and $C_{st}^{pq_t}$ equal to 0. $C_{tt}^{pq_t}$ is equal to dissimilarity penalty $L(p, q_t)$ and $C_{ts}^{pq_t}$ is equal to the occlusion penalty. Note that the value of K^{occ} has to be greater or equal to $L(p, q_t)$ in order to respect the constraint given by (13). To that end, we truncate $L(p, q_t)$ to the value of K^{occ} .

The smoothness of the result is ensured by term E^s of the energy given in (7). We use the following definition:

$$E_i^s(d_i) = \sum_{(u,v) \in \Psi_i} \|d_i(u) - d_i(v)\|, \quad (14)$$

where Ψ_i is the set of neighbour pixels pairs in image i . We use two implementations of the smoothness constraint. In the first one Ψ_i contains only horizontal neighbours in order to obtain a 1D smoothness constraint. Such a constraint makes for an independent selection of epipolar lines and then a parallel implementation. The second one uses a 2D smoothness constraint and includes both horizontal and vertical neighbours.

5. Results

To compare our methods, first between them and secondly with other existing ones, we use two sets of 8 images. The first one is a set of images of a virtual scene, which allows us to compare results against ground truth. The second one is a set of photographs taken at Palais du Tau in Reims [18]. The dimensions of images in both sets are 512×384 pixels. Figure 5 shows one image of each set. The photograph has homogeneous colors whereas the virtual scene has various colors. This allows us to test our methods in both cases.

We compare three pairs of methods. The first pair is composed of correlation-based methods. One uses the cost of (2) and the other is our own correlation-based method.



FIGURE 5: A photography shot at Palais du Tau (a) and a virtual scene (b).

TABLE 1: Errors with ground truth on the virtual scene plus computation times on both Palais du Tau and virtual sets.

Method	Occlusion	Error	Times (s)	
			Tau	Virtual
Correlation	without	0.644	0.23	0.25
	with	0.630	1.25	1.24
1D smoothness	without	0.718	6.88	6.02
	with	0.572	17.89	12.39
2D smoothness	without	0.692	12.02	8.98
	with	0.543	38.13	18.36

The second and third pairs are graph-cuts based methods. One with a smoothness constraint along epipolar lines, and one with a 2D smoothness constraint. Each pair is composed of a method using our energy function and one using a standard energy of the form

$$E_0(d_0) = E_0^d(d_0) + \chi E_0^s(d_0). \quad (15)$$

Using the ground truth of the virtual scene, we give the error rate corresponding to each method in Table 1. This error results from the absolute differences between real disparities and our disparities summed on all pixels and divided by the number of pixels. The result is then the average disparity error per pixel. Our local methods have the highest error rate, and our global method has the smallest error rate. For each category, the error rate of the method using our occlusion approach is always lower than the method which does not use it. We observe that the error rate of the correlation-based method without occlusion handling is lower than the one of the method using 1D smoothness constraint. We think that is due to the fact that our virtual images contain no noise and no specular light at all. That is the reason why the correlation-based method gives particularly good results in this case. Therefore, the comparison between correlation-based methods and energy minimization based methods is meaningless.

Table 1 also gives computation times on both sets of images. We used an Intel Core 2 Duo CPU E4700 and 2 Go of memory. Both correlation-based methods are implemented using CUDA on an NVIDIA Quadro FX 3700 graphic card. Times include the computation of the whole set of the N disparity maps. Globally, methods using the occlusion approach are slower than the other methods. This is due to the fact that they have more possibilities to take into

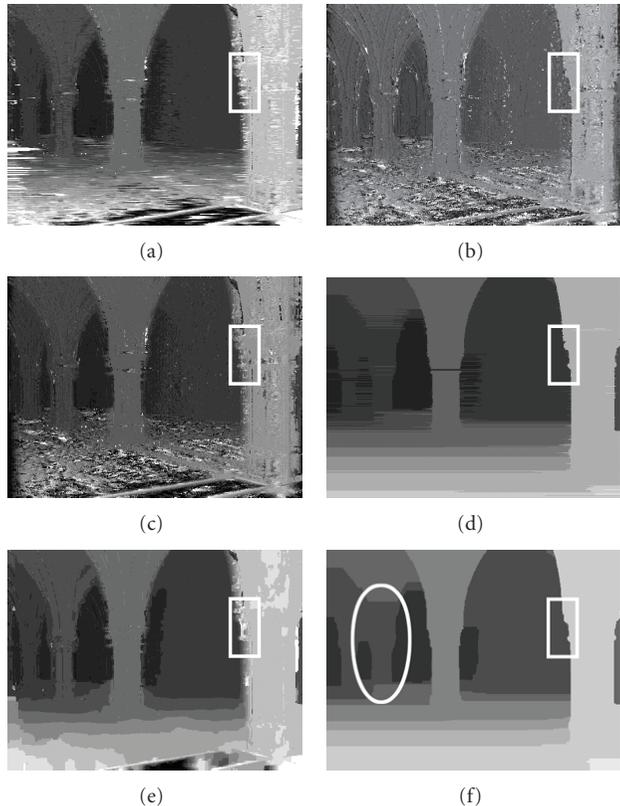


FIGURE 6: Disparity maps obtained using methods based on: correlation (a),(b), horizontal smooth constraint (c),(d), and 2D smooth constraint (e),(f). Images (a),(c),(e) and (b),(d),(f) are, respectively, computed without and with our occlusion approach.

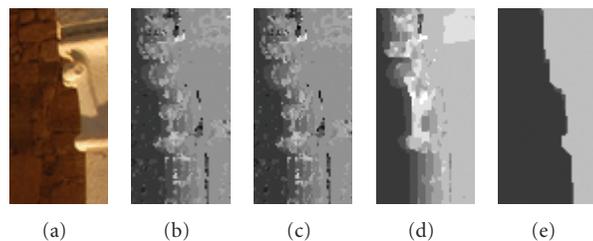


FIGURE 7: Extracts of Figures 5, 6(a), 6(b), 6(e), and 6(f).

account. Indeed, the correlation-based method has more tests to carry out, and the graph-cuts based method has a more complicated graph structure, that is slower to solve.

Figure 6 shows results obtained without and with occlusion handling ($K^{\text{occ}} = 100$). Figure 6(a) illustrates the standard method; whereas Figure 6(b) corresponds to the method we have presented. We can see in these images that our method has precisely detected occlusions on boundaries of the columns. However, areas without depth discontinuities like the background contain errors. We think that it is due to the noise sensitivity of our occlusion detection. The first three extracts given in Figures 7(a), 7(b), and 7(c) show the details of these results on the front column. We observe that our correlation-based method has accurately defined

discontinuities fitting the real boundaries of the column, but disparities in occluded areas contain errors (noise in Figure 7(c)), since the method has difficulties to find them in such areas. Finally, our approach is not well suited to the principle of correlation-based methods. In fact, in such methods the selection of disparities is independent for each pixel, and the consistency from one disparity map to another cannot be ensured.

On the other hand, graph-cuts based methods allow the symmetrical minimization of energy, ensuring a strong consistency. Figures 6(c) and 6(d) show results with the 1D smoothness constraint, and Figures 6(e) and 6(f) show results with the 2D smoothness constraint. Again, the method using our occlusion approach has very accurate depth discontinuities at object boundaries. In our 1D smoothness method (Figure 6(d)), the horizontal line effect is more visible than in the classical method (Figure 6(c)). This is due to the fact that our occlusion approach penalizes any disparity variation, because it is detected as an occlusion. Our method tends to add a strong smoothness constraint along epipolar lines. The 2D smoothness constraint allows compensating for this artifact. The images in Figures 7(d) and 7(e) show the front column obtained with these methods. We notice that without occlusion handling the method cannot find the disparities of some pixels, which are not visible in all images. On the other hand, it accurately detects the occlusion using our occlusion approach. However, we observe in Figure 6(f) that the column in the background (in a white ellipse) is not well defined. This is due to the principle of plane-sweeping algorithms and to the fact that this column is actually between two planes. The reader can refer to [14] for a presentation of our refinement step, which solves this problem.

6. Conclusion

We have introduced a new approach in order to handle occlusions of a scene in a multiview context. As a proof of the relevance of this new detection rule, we have presented two methods with the particularity of handling objects boundaries very accurately. Even if these methods can handle two-view stereovision, they are designed for the multiview context with any number of views. The results we obtain show that our occlusion approach succeeds in detecting objects boundaries to the detriment of computation times, and can still create disparities even for pixels that are not visible in all views. Moreover, used on symmetrical energy minimization-based methods, our approach ensures a geometric consistency, which is crucial for autostereoscopic displays. However, computation time is the main problem of our methods. That is the reason why our objective is to find a means to minimize energy faster. One idea is the GPU implementation of the graph cuts. Some work has already been done in this domain [19, 20] but we are not using it for the moment since it induces a lot of constraints on the graph structure and must be adapted to our energy. Another possibility that we are working on is to reduce the number of nodes used in our graph, in order to simplify the maximum flow problem.

Acknowledgments

The work presented in this paper was supported in part by the “Agence Nationale de la Recherche” as part of the CamRelief project. This project is a collaboration between the University of Reims Champagne-Ardenne and 3DTV Solutions. The authors wish to thank Didier Debons, Michel Frichet, and Florence Debons for their contribution to the project.

References

- [1] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *International Journal of Computer Vision*, vol. 47, no. 1–3, pp. 7–42, 2002.
- [2] A. Klaus, M. Sormann, and K. Karner, “Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure,” in *Proceedings of the 18th International Conference on Pattern Recognition (ICPR '06)*, vol. 3, pp. 15–18, Hong Kong, August 2006.
- [3] Z.-F. Wang and Z.-G. Zheng, “A region based stereo matching algorithm using cooperative optimization,” in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, Anchorage, Alaska, USA, June 2008.
- [4] G. Zhang, J. Jia, T.-T. Wong, and H. Bao, “Recovering consistent video depth maps via bundle optimization,” in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, Anchorage, Alaska, USA, June 2008.
- [5] Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér, “Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 492–504, 2009.
- [6] G. Egnal and R. P. Wildes, “Detecting binocular half-occlusions: empirical comparisons of five approaches,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1127–1133, 2002.
- [7] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert, “Dense disparity map estimation respecting image discontinuities: a PDE and scale-space based approach,” *Journal of Visual Communication and Image Representation*, vol. 13, no. 1–2, pp. 3–21, 2002.
- [8] M. Proesmans, L. van Gool, E. Pauwels, and A. Oosterlinck, “Determination of optical flow and its discontinuities using non-linear diffusion,” in *Proceedings of the 3rd European Conference on Computer Vision (ECCV '94)*, vol. 2, pp. 295–304, Springer, Secaucus, NJ, USA, 1994.
- [9] L. Alvarez, R. Deriche, T. Papadopoulos, and J. Sánchez, “Symmetrical dense optical flow estimation with occlusions detection,” in *Proceedings of the 7th European Conference on Computer Vision-Part I (ECCV '02)*, pp. 721–735, Springer, London, UK, 2002.
- [10] C. Strecha and L. Van Gool, “PDE-based multi-view depth estimation,” in *Proceedings of the 1st International Symposium on 3D Data Processing Visualization and Transmission (3DPVT '02)*, pp. 416–425, 2002.
- [11] S. Ince and J. Konrad, “Geometry-based estimation of occlusions from video frame pairs,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*

- (*ICASSP '05*), vol. 2, pp. 933–936, Philadelphia, Pa, USA, March 2005.
- [12] P.-M. Jodoin, C. Rosenberger, and M. Mignotte, “Detecting half-occlusion with a fast region-based fusion procedure,” in *Proceedings of the British Machine Vision Conference*, pp. 417–426, 2006.
 - [13] S. Ince and J. Konrad, “Occlusion-aware optical flow estimation,” *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1443–1451, 2008.
 - [14] C. Niquin, S. Prévost, and Y. Remion, “Accurate multi-view depth reconstruction with occlusions handling,” in *Proceedings of the 3rd 3DTV-Conference on the True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON '09)*, Potsdam, Germany, May 2009.
 - [15] J. Woetzel and R. Koch, “Real-time multi-stereo depth estimation on GPU with approximative discontinuity handling,” in *Proceedings of the 1st European Conference on Visual Media Production (CVMP '04)*, pp. 245–254, London, UK, March 2004.
 - [16] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.
 - [17] V. Kolmogorov and R. Zabih, “What energy functions can be minimized via graph cuts?” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, 2004.
 - [18] <http://palais-tau.monuments-nationaux.fr>.
 - [19] A. Bhusnurmath and C. J. Taylor, “Graph cuts via ℓ_1 norm minimization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1866–1871, 2008.
 - [20] V. Vineet and P. J. Narayanan, “CUBA cuts: fast graph cuts on the GPU,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '08)*, pp. 1–8, Anchorage, Alaska, USA, June 2008.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

