REVISITING INTRINSIC CURVES FOR EFFICIENT DENSE STEREO MATCHING

M. Shahbazi [a, *], G. Sohn [b], J. Théau [a], P. Ménard [c]

[a] Dept. of Applied Geomatics, Université de Sherbrooke, Boul. de l'Université, Sherbrooke, Québec, Canada - (mozhdeh.shahbazi, jerome.theau)@usherbrooke.ca
[b] Dept. of Geomatics Engineering, York University, Keele Street, Toronto, Ontario, Canada - gsohn@yorku.ca
[c] Centre de géomatique du Québec, Saguenay, Québec, Canada - pmenard@cgq.qc.ca

Commission III, WG III/1

KEY WORDS: Matching, Intrinsic Curves, Search Space Reduction, Occlusion, Energy, Optimization

ABSTRACT:

Dense stereo matching is one of the fundamental and active areas of photogrammetry. The increasing image resolution of digital cameras as well as the growing interest in unconventional imaging, e.g. unmanned aerial imagery, has exposed stereo image pairs to serious occlusion, noise and matching ambiguity. This has also resulted in an increase in the range of disparity values that should be considered for matching. Therefore, conventional methods of dense matching need to be revised to achieve higher levels of efficiency and accuracy. In this paper, we present an algorithm that uses the concepts of intrinsic curves to propose sparse disparity hypotheses for each pixel. Then, the hypotheses are propagated to adjoining pixels by label-set enlargement based on the proximity in the space of intrinsic curves. The same concepts are applied to model occlusions explicitly via a regularization term in the energy function. Finally, a global optimization stage is performed using belief-propagation to assign one of the disparity hypotheses to each pixel. By searching only through a small fraction of the whole disparity search space and handling occlusions and ambiguities, the proposed framework could achieve high levels of accuracy and efficiency.

1. INTRODUCTION

Dense stereo matching has always been one of the fundamental and active areas of photogrammetry and computer vision. In addition to three-dimensional (3D) scene reconstruction, several other applications such as view synthesis, image-based rendering and robotics benefit from the results of dense matching. Generally, any dense stereo matching technique can be described by the following three components: matching cost computation, cost aggregation and disparity computation (Scharstein and Szeliski, 2002). Disparity refinement, which is basically enhancement of the generated depth field, can also be considered as the fourth step of the dense matching.

Most of the dense matching techniques, which are discussed in the next section, still have considerable computational complexity regarding the size of the disparity search space. The techniques proposed to deal with such complexities either evaluate the complete disparity space implicitly or require successive correspondence search over the full or limited range of disparities. In addition, the sensitivity of matching techniques to occlusion, noise and matching ambiguity at poorly-textured regions is undeniable. These necessitate enforcing constraints such as ordering and uniqueness that are not always efficient/true for all types of scenes.

In this paper, we present a matching algorithm that is generally based on the concepts of intrinsic curves (Tomasi and Manduchi, 1998). This matching strategy avoids exhaustive disparity search space exploration by proposing sparse disparity hypotheses for each pixel. Then, the hypotheses are propagated to adjoining pixels by label-set enlargement based on the proximity in the space of intrinsic curves in order to avoid gaps created due to noise. The same concepts are applied to model occlusions explicitly via a regularization term in the energy function. Finally, a global optimization stage is performed using belief-propagation to assign one of the disparity hypotheses to each pixel.

The rest of the paper is organized as follows. First, a brief review of literature in dense stereo matching is presented. Then, the details of the proposed technique are presented in sections 3 and 4. The experiments performed to evaluate our method are discussed in section 5 and, finally, the conclusions are mentioned in section 6.

2. RELATED WORK

The techniques of dense matching can be classified into two categories of local and global methods. Local methods construct a cubic cost volume $C(x,y,d)$, which represents the cost associated with matching a pixel (x,y) in the left (reference) image to the corresponding pixel in the other stereo image (right image) at a disparity value d belonging to the full disparity search range. This matching cost is usually quantified by a per-pixel dissimilarity measure. Then, the disparity map can be determined by finding the minimum cost at each pixel as $\hat{d}(x,y) = \arg\min_d C(x,y,d)$. However, such results are highly noisy because the solution is not regularized. To regularize the solution one can aggregate the costs over a support region (local window) and find the disparity with the lowest aggregated cost. Basically, the window-based cost aggregation means filtering the (x,y) dimensions of the cost volume (Hosni et al., 2013). Thus, the aggregated cost over a window W can be achieved as $C_S(x,y,d) = \sum_{(u,v)\in W} \omega_{(u,v)} C(x,y,d)$, where ω is the weight of a pixel in the support area (Gurbuz et al., 2015). For instance, in the sum-of-squared-differences (SSD) algorithm, the support

* Corresponding author

aggregation is done by summing matching costs over a squared window surrounding each point assuming a constant disparity, i.e. ω=1. In some algorithms, the aggregation is performed implicitly by computing a window-based matching cost such as normalized-cross-correlation (NCC). Although these window-based aggregations yield smoother results, they ignore disparity discontinuities since the windows are not aligned with image edges. In order to preserve depth discontinuities, edge-aware weighted filters, e.g. geodesic distance, bilateral and guided filtering, can be used at the cost of noticeably higher computational complexity (Yoon and Kweon, 2006; Hosni et al., 2009; Hosni et al., 2013). The main drawback of local techniques is that they require evaluating the full disparity space image (DSI). That is the matching costs should be computed and aggregated at each pixel for all possible disparities (Sinha et al., 2014). They also depend largely on the choice of the window size (W). While a small window is preferred to avoid over-smoothing and to increase computational efficiency, a large window is required in areas of low texture to decrease matching ambiguity. These facts results in a trade-off between lower success rate using small windows and border bleeding artefacts using large windows (Geiger et al., 2011). Therefore, when the local characteristics of the pixels are similar, considerable ambiguity is involved in finding their correspondences without global reasoning.

In global methods, stereo matching is formulated as a pixel labelling problem, where the inputs are a set of pixels and a set of labels (i.e. potential disparities). From the probabilistic point of view, this can be treated as an inference problem using Bayesian approaches. That is inferring the disparity map given the likelihood based on image observations and the prior based on the assumptions about the scene structure (e.g. disparity smoothness) (Sun et al., 2013). Bayesian approaches can be divided into two categories based on path-finding and Markov random fields (MRF). In path-finding approaches, ordering and uniqueness constraints are the priors that are used to solve scanline matching in terms of finding the shortest path to go from the beginning to the end of the corresponding scanlines (epipolar lines) in the matrix of their pair-wise matching costs. The most popular solution used in the literature to find such minimum-cost path is dynamic programming (DP) (Cox et al., 1996). The main drawback of these techniques is that the dependence between scanlines is either totally ignored or partially considered via inter-scanline vertical edges (Ohta et al., 1985). Without the smoothness assumption between scanlines, the results of path-finding approaches often suffer from streaking effect (Zitnick and Kanade, 2000). As a powerful alternative, the labelling problem can be modelled as a Markov random field (MRF) since a particular pixel label depends only on the labels of its neighbours (Boykov et al., 1998). Such problem can then be solved in an energy minimization framework, where a global energy function penalises intensity dissimilarities between the corresponding pixels and discontinuities between the neighbouring pixels in the disparity map (Scharstein and Szeliski, 2002). In other words, the maximum a posterior (MAP) estimate of the disparity map can be achieved by minimizing this energy function. However, minimizing a global MRF-based energy function is generally NP-hard. Therefore, a variety of approximation algorithms have been proposed that apply graph cuts or belief propagation for inference (Sun et al., 2003; Boykov et al., 2001). Semi-global matching (SGM) is also a technique that approximates the global MRF inference by aggregating cost functions along several (usually eight or sixteen) local paths in the image and computes the disparity with a winner-take-all mechanism (Hirschmüller, 2008).

A drawback of global methods is their computational complexity, which does not scale well to large label spaces (large DSI). Several solutions exist for limiting the disparity search range. The simplest way is to select only the best disparities for each pixel that correspond to the highest matching scores. Another way is the hierarchical approach, in which a Gaussian pyramid of the original images is constructed and disparities are computed at each level. Then, disparity results at coarser levels are used to reduce the disparity range at finer levels. Local, fast window-based techniques can also be used before applying a complex global technique to reduce the matching ambiguity. A more recent category of techniques for disparity search space reduction is based on assuming planar hypotheses for specific regions of the images. For instance, Libelas method builds a prior on the disparities by forming a triangulation on sparsely matched key-points (Geiger et al., 2011). PatchMatch stereo also finds correspondences between small patches (segments) of the images by iteratively propagating disparities from initial seeds to their neighbours (Bleyer et al., 2011). In LPS method, local slanted plan hypotheses, which are derived from initial sparse feature correspondences, are used to propose disparity hypotheses (Sinha et al., 2014). Except for the best-candidate based method, the rest of these techniques require either exhaustive DSI computation or successive matching to find the appropriate priors.

Another drawback of global methods is their tendency to fail in occluded regions (Mozerov et al., 2015). Generally, uniqueness and ordering constraints are used to handle occlusion. Uniqueness is usually enforced by left-right cross-checking of disparity maps. The ordering constraint, which is mostly satisfied in DP-based methods, requires that the relative ordering of pixels remains the same on the corresponding scanlines. However, such assumption is not always true, especially in scenes containing narrow foreground objects (Scharstein and Szeliski, 2002).

3. ORIGINAL CONCEPTS OF INTRINSIC CURVES

3.1 Definition

In this study, we revisit the concept of intrinsic curves proposed by Tomasi and Manduchi (1998) to develop a new dense matching approach. Intrinsic curves are the multidimensional representations of the paths that the image descriptors follow as a scanline is traversed from left to right. Assume the intensity and its derivative (gradient) as two image descriptors. Scanlines can then be considered as one-dimensional (1D) signals of intensity $I(x^l)$ at every location x^l on the left scanline and $r(x^r)$ at every location x^r on the right scanline (Figure 1a)[1]. Respectively, their derivatives are $I'(x^l)$ and $r'(x^r)$. If we plot $I'(x^l)$ versus $I(x^l)$, then we lose the spatial track of x^l (Figure 1b), i.e. the new representation of the left scanline would be $C^l=I'(I)$. In this new representation, if $I(x)$ is replaced by a displaced replica $I(x^l+d)$, the curve C^l of Figure 1b remains the same. Due to this invariance to displacements (disparities), these curves are called intrinsic curves. Then, the problem of image matching would be converted to the problem of curve matching.

It should be noted that in order to have a continuous/smooth representation of the curves, the scanline intensity signals are modelled with piece-wise cubic splines with regard to x.

[1] The stereo images belong to Middlebury Stereo Datasets; see Section 5.

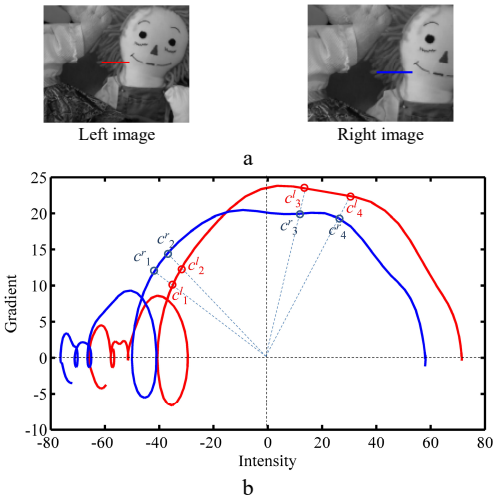

Figure1. (a) One part of left and right corresponding scanlines; (b) Intrinsic curves of the left and right scanlines (C^l, C^r) in red and blue, respectively.

3.2 Matching Cost Computation

If the only difference between two scanlines is the geometric disparity, the two curves would coincide everywhere. However, the images are usually corrupted by noise and photometric transformations. Therefore, there is a non-constant variation between the two curves. According to Tomasi and Manduchi (1998), zero-mean low-pass filtering of the images may remove the noise and brightness bias between images completely. Therefore, contrast difference is the only remaining difference between scanlines, which causes the right intrinsic curve C^r be an expanded/contracted form of the left intrinsic curve C^l from the origin. This assumption suggests a radial metric for finding candidate matches based on intrinsic curves; i.e. two points might be corresponding if they are collinear with the origin in the two-dimensional space of the curves, e.g. points $\boldsymbol{c}^l_1$ and $\boldsymbol{c}^r_1$ in Figure 1b. In this sense, their radial distance in the full space of curves shows the degree of their similarity and measures the matching cots.

3.3 Local Aggregation

The original study suggests aggregating the candidate matches into candidate matching segments; i.e. a candidate match and its close neighbours on the intrinsic curves belong to a matching segment, and these matches are either all wrong or right. For instance, candidate correspondences ($\boldsymbol{c}^l_1$, $\boldsymbol{c}^r_1$) and ($\boldsymbol{c}^l_2$, $\boldsymbol{c}^r_2$) belong to the same segment since the arc-length s_l between $\boldsymbol{c}^l_1$ and $\boldsymbol{c}^l_2$ is smaller than a threshold and is also close to the arc-length s_r between $\boldsymbol{c}^r_1$ and $\boldsymbol{c}^r_2$. The matching cost of each segment is defined as the sum of the matching costs of its correspondences.

3.4 Disparity Computation

Because of the aggregation step, there are fewer candidate segments than candidate matching points; therefore, the search space for matching is reduced. Constraints of uniqueness and ordering are applied to the candidate segments to solve the matching with a path-finding approach. That is, two candidate segments, e.g. $S_1=\{(\boldsymbol{c}^l_1,\boldsymbol{c}^r_1),(\boldsymbol{c}^l_2,\boldsymbol{c}^r_2)\}$ and $S_2=\{(\boldsymbol{c}^l_3,\boldsymbol{c}^r_3),(\boldsymbol{c}^l_4,\boldsymbol{c}^r_4)\}$, can follow each other only if there is no overlap between them (uniqueness) and one is the successor of the other (ordering). Based on this concept, a graph is formed in which the candidate segments are the nodes, and segments that can follow each other are linked with edges. The edges are weighted based on the aggregated matching cost of the source nodes. Then, an application of a shortest-path-finding algorithm produces the minimum-cost path through the scanline, that is, the best matching segments.

3.5 Shortcomings of the Original Concepts

Although low-pass filtering of a signal reduces its noise level and subtracting the mean from two signals decreases their shift (bias), there usually remains some local shift between the signals. In other words, assuming an affine photometric distortion model is locally possible, but not globally. As a result, the radial metric (phase similarity) is not applicable everywhere. For example, at the stereo pair of Figure 1a, only for 35% of the points, the ground-truth match exists among the hypothesized candidate matches.

In addition, computing matching cost as the radial distance between two points is not enough to resolve the matching ambiguities since this distance is only a pixel-wise measure, similar to squared intensity difference (SD).

Another shortcoming arises from the aggregation step. While segmenting candidate matches based on the arc-length proximity seems like a brilliant alternative for window-based cost aggregation, it does not provide dense matching and struggles for finding correct matches at poorly textured image areas. To understand this more clearly, a part of a scanline with both poor and high-variance textures is shown in Figure 2. Sampling the intensity signal of this scanline based on a constant x (1 pixel) results in a uniform grid through the scanline (Figure 2a). However, sampling based on a constant arc-length (9 gray values) through the intrinsic curve results in a non-uniform grid through the scanline (Figure 2b). This non-uniform grid is denser at busy areas of the image since the arc-lengths are longer where high intensity changes happen. As a result, at these areas candidate matches may not be close enough (in terms of arc-length) to form candidate segments; therefore, they will be left unmatched. On the other hand, at flat areas of the image, the arc-lengths are shorter. Thus, these areas produce crowded candidate segments with high ambiguity that result in wrong matches.

4. PROPOSED DENSE MATCHING ALGORITHM

In contrast with the original concept, we do not want to lose the full track of disparity. However, we are looking for a solution to maximize the benefits from the concepts of intrinsic curves in order to first, reduce the disparity search space without requiring successive matching or additional search, and, second, to handle the poorly-textured areas (ambiguity) and occluded areas (occlusion) efficiently.

4.1 Hypothesis Generation

We establish the hypotheses for pixels belonging to a small neighbourhood (N_i) by assuming that the photometric transformation between corresponding scanlines can be locally modelled with an affine transformation as follows:

$$r(x^l + d) = \alpha_i l(x^l) + \beta_i \quad , x^l \in N_i \tag{1}$$

where the gain parameter α_i and the shift parameter β_i represent the difference in contrast and brightness between the scanlines

in the local neighborhood N_i. This assumption necessitates zero-mean low-pass filtering of the image, e.g. with a Gaussian filter with kernel size 3 and standard deviation of 1.2, in order to avoid an additional noise term in Equation (1). In fact, the noise term can be considered small enough to be independent of x and to be integrated to the brightness bias term.

If we consider the parametric representation of the left intrinsic curves by a pair of functions as $C^l(x) = (l(x), l'(x))$, then the parametric form of the right curve can be estimated as $C^r(x^l + d) = (\alpha_i l(x^l) + \beta_i, \alpha_i l'(x^l))$. This proves that, at corresponding points, C^l and C^r cannot be radial from the centre; however, the tangents to the curves, $\boldsymbol{t^l} = (l', l'')$ and $\boldsymbol{t^r} = (r', r'')$, have the same orientations (θ^l, θ^r).

$$\theta^l(x^l) = \tan^{-1}\left(\frac{l''}{l'}\right)$$
$$\theta^r(x^l + d) = \tan^{-1}\left(\frac{r''}{r'}\right) = \tan^{-1}\left(\frac{\alpha_i l''}{\alpha_i l'}\right) \quad (2)$$

Therefore, at any pixel $\boldsymbol{p}$ at coordinates (x_p, y_p) on the left image, the set of disparity candidates D_p can be defined as follows:

$$D_p = \{d \in D \mid abs(\theta^l(x_p) - \theta^r(x_p + d)) < T\} \quad (3)$$

where D is the whole range of possible disparities.

Although this assumption is generally true for small thresholds T, it may cause problems at texture-less areas of the image, where intensity changes are either very low or zero. Because, in these cases, even small values of remaining noise on the scanlines are large enough, compared to the values of (l', l''), to make noticeable differences between θ^l and θ^r. To avoid such situations, we can increase the threshold T. In addition, a filtering step is applied to fill in any gaps in the disparity hypotheses which may exist because of noise. This filtering step is similar to the idea proposed by Veksler (2006), however with different definition and measures. Let define set P_d as $P_d = \{\boldsymbol{p} \mid d \in D_p\}$. In other words, P_d is the set of all the pixels for which the disparity d is selected as a candidate disparity. Now, P_d can be extended (enlarged) to all the pixels which are close to pixels of set P_d on the intrinsic curves. That is:

$$P_d^{extended} = \{\boldsymbol{p_i} \mid \varphi(\boldsymbol{p_i}, \boldsymbol{p_j}) < \sigma \quad \text{for some } \boldsymbol{p_j} \in P_d\} \quad (4)$$

where $\varphi(\boldsymbol{p_i}, \boldsymbol{p_j})$ is the curve arc-length measured as:

$$\varphi(\boldsymbol{p_i}, \boldsymbol{p_j}) = \int_{xp_i}^{x_{pj}} \left((l'(x))^2 + (l''(x))^2\right)^{1/2} dx \quad (5)$$

It is noteworthy that this measure to extend the disparity candidates considers the fact that close pixels (in spatial space) that have similar intensities and similar intensity changes (therefore, are close in the curve space) are more probable to have similar disparities.

4.2 Occlusions

Using intrinsic curves, occlusions stand out as pieces of one curve, usually in form of loops, which remain unmatched in the other curve. Figure 3a and Figure 3b show two scanlines and their corresponding intrinsic curves. The stars on the curves show the matched pixels and the circles on the left curve show the occluded pixels.

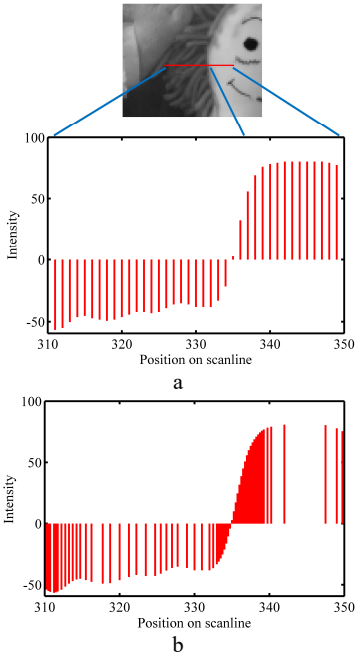

Figure 2. (a) Uniform sampling of a scanline based on position; (b) Non-uniform sampling based on arc-length.

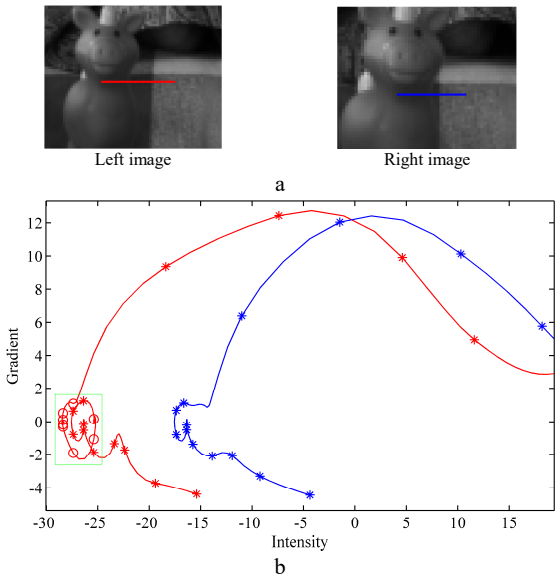

Figure 3. (a) One part of left and right corresponding scanlines with partial occlusions; (b) Intrinsic curves demonstrating matched pixels (shown by stars) and occluded pixels (shown by circles in the green frame).

As it can be noticed, occlusions happen at left arcs whose curvature changes are not similar to the right arcs. In other words, the curvature of the left curve changes the same way as the curvature of the right curve between any two non-occluded corresponding points; i.e. the curve remains either concave or convex in both curves. To formulate this concept

mathematically, consider a pixels x^l_i on the left scanline and its adjacent neighbours x^l_i+1 and x^l_i-1. Their correspondences on the right scanline are denoted as x^r_i, x^r_i+1 and x^r_i-1. Respectively, the points on the intrinsic curves associated with these pixels are denoted as $\boldsymbol{c}^l_i$, $\boldsymbol{c}^l_{i+1}$, $\boldsymbol{c}^l_{i-1}$, $\boldsymbol{c}^r_i$, $\boldsymbol{c}^r_{i+1}$ and $\boldsymbol{c}^r_{i-1}$ (Figure 4). Let's assume that ${}^l\kappa_i^{i+1}$ contains the curvature values of the left curve between $\boldsymbol{c}^l_i$ and $\boldsymbol{c}^l_{i+1}$. Similar definition applies to ${}^l\kappa_{i-1}^{i}$, ${}^r\kappa_i^{i+1}$ and ${}^r\kappa_{i-1}^{i}$. Therefore, the similarity of curvature changes around pixels x^l_i and x^r_i can be measured as follows.

$$\Theta(x_i^l, x_i^r) = \min_{j\in\{i,i-1\}} \left(\left| sign({}^l\kappa_j^{j+1}) - sign({}^r\kappa_j^{j+1}) \right| \right) \tag{6}$$

If x^l_i is non-occluded and is surrounded by two non-occluded pixels x^l_i+1 and x^l_i-1, then $sign({}^l\kappa_i^{i+1})$ and $sign({}^r\kappa_i^{i+1})$ are the same; besides, $sign({}^l\kappa_{i-1}^{i})$ and $sign({}^r\kappa_{i-1}^{i})$ are equal too, which means $\Theta(x_i^l, x_i^r) = 0$. Alternatively, if it is surrounded by one occluded pixel at one side, for example at x^l_i+1, and by one non-occluded pixel at the other side (x^l_i-1), then only $sign({}^l\kappa_{i-1}^{i})$ and $sign({}^r\kappa_{i-1}^{i})$ are equal; however, $sign({}^l\kappa_i^{i+1})$ and $sign({}^r\kappa_i^{i+1})$ are different. This shows an occurrence of occlusion. This concept is used later in defining the global energy function for matching.

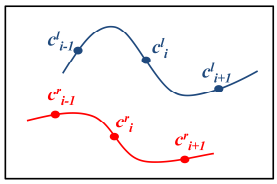

Figure 4. Example of intrinsic curves to measure local curvature similarity for occlusion detection

It should be noted that occlusions are also implicitly considered in the stage of hypothesis generation. That is, the pixels with empty candidate disparity sets are automatically recognized as occluded pixels and will be left unmatched.

4.3 Global Energy Function

Consider the reference image as a set of pixels P with observed intensities I_p for each $\boldsymbol{p} \in P$. Therefore, a pixel $\boldsymbol{p}$ is representative of image observations at pixel $\boldsymbol{p}$. The pixel $\boldsymbol{p}$ has coordinates (x_p,y_p) and its four immediate neighbours $\{(x_p-1,y_p), (x_p+1,y_p), (x_p,y_p-1), (x_p,y_p+1)\}$ form a neighbourhood set N_p. The set of all candidate disparities D is defined as $D = \bigcup_{\boldsymbol{p}\in P} D_p$, where D_p is the set of disparity hypotheses for pixel p as defined in Section 4.1.

In global algorithm, the goal of matching is to compute, for each pixel $\boldsymbol{p}$, an optimal label d_p as the disparity at this pixel, so that the following energy function is minimized:

$$E(P,D) = \sum_{\boldsymbol{p}\in P} \left\{ \begin{matrix} E_{data}(\boldsymbol{p}, d_p) + E_{occ}(\boldsymbol{p}, d_p) \\ + \sum_{\boldsymbol{q}\in N_p} E_{smooth}(d_p, d_q) \end{matrix} \right\} \tag{7}$$

In the following subsections, the data term E_{data} and smoothness term E_{smooth}, the occlusion term E_{occ} are defined. Later in Section 4.4, the belief propagation strategy to minimize this energy function is explained as well.

4.3.1 Data Term: The data term encodes the intensity similarity (photometric consistency) of pixel correspondences for hypothesized disparities:

$$E_{data}(\boldsymbol{p}, d_p) = \rho_{data}(F(\boldsymbol{p}, d_p)) \tag{8}$$

where $F(\boldsymbol{p}, d_p)$ is the cost of matching pixel $\boldsymbol{p}$ to the right image with disparity d_p. In this study, we measure F based on non-parametric census transform (Zabih and Woodfill, 1994). The function ρ_{data} is a truncated L_1 norm function proposed by (Sun et al., 2003) that is robust to noise:

$$\rho_{data}(t) = -\ln\left((1-e_d)\exp\left(-\frac{|t|}{\sigma_d} \right) + e_d \right) \tag{9}$$

where e_d and σ_d are parameters controlling the shape of the function.

4.3.2 Smoothness Term: The smoothness term encodes the piecewise smoothness prior on the disparities of every pixel $\boldsymbol{p}$ and its immediate neighbours $\boldsymbol{q}$.

$$E_{smooth}(d_p, d_q) = \rho_{smooth}(d_p - d_q) \tag{10}$$

where the function ρ_{smooth} is also a robust function defined similar to Equation (9) with parameters e_s and σ_s.

$$\rho_{smooth}(t) = -\ln\left((1-e_s)\exp\left(-\frac{|t|}{\sigma_s} \right) + e_s \right) \tag{11}$$

The function ρ_{smooth} has the form of a potential function of Total Variance (TV), which has the discontinuity preserving property needed for a proper smoothness term (Sun et al., 2003). The parameter e_s controls the upper bound (truncation) of the model and the parameter σ_s defines the sharpness (Figure 5).

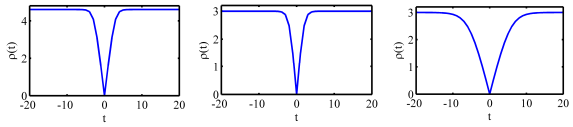

Figure 5. The robust function $\rho(t)$ with different parameters. (a) e=0.01, σ=0.6; (b) e=0.05, σ=0.6; (c) e=0.05, σ=1.8.

4.3.3 Occlusion Term: In addition, we want to include the occlusion term for encoding the occlusion clues as explained in Section 4.2. Therefore, we integrate such occlusion assumption into the basic energy function using a soft constraint as E_{occ}.

$$E_{occ}(\boldsymbol{p}, d_p) = \Theta(x_p, x_p + d_p) \tag{12}$$

where the function Θ is defined as in Equation (6). This energy term imposes a penalty for a pixel $\boldsymbol{p}$ being occluded on the right image assuming a disparity d_p.

4.4 Approximating Inference by Belief Propagation

From probabilistic point of view, the energy is equal, up to a constant, to the negative log posterior. Therefore, minimizing the energy function $E(P,D)$ is equivalent to maximizing a posterior probability $p(D|P)$; that is the disparity map given image observations (Equation (13)). In fact, in a Markov random field network, pixels P can be considered as observable nodes of the graph, and disparity hypotheses D (or labels) can be considered as random variables (hidden nodes). There is a link (undirected edge) between any pixel $\boldsymbol{p}$ and all the labels in its candidate disparity set D_p and also between any directly neighbouring pixels:

$$posterior \propto \prod_{\boldsymbol{p}\in P}\left(\gamma(\boldsymbol{p},d_p)\prod_{\boldsymbol{q}\in N_p}\eta(d_p,d_q)\right) \quad (13)$$

where γ is the unary potential,

$$\gamma(\boldsymbol{p},d_p)=\exp\left(-\rho_{data}(F(\boldsymbol{p},d_p)-\Theta(x_p,x_p+d_p)\right) \quad (14)$$

and η is the interaction or binary potential.

$$\eta(d_p,d_q)=\exp(-\rho_{smooth}(d_p-d_q)) \quad (15)$$

To maximize this posterior, loopy belief propagation (LBP) technique is used (Sun et al., 2003). There are several algorithms to perform LBP, from which the max-product algorithm is applied here that maximizes the joint posterior. Note that this would be equivalent to the min-sum algorithm, if the negative log probabilities were used instead of the potentials. Max-product is basically a message-sending algorithm where every observable node $\boldsymbol{p}$ sends a message ($msg_{p\to q}(d)$) to node $\boldsymbol{q}$ in its neighbourhood about the amount of its belief that node $\boldsymbol{q}$ has disparity value d:

$$msg_{\boldsymbol{p}\to\boldsymbol{q}}(d)=\max_{l\in D_p}\left(\gamma(\boldsymbol{p},l)\eta(d,l)\prod_{\boldsymbol{s}\in N_p\setminus\boldsymbol{q}}msg_{\boldsymbol{s}\to\boldsymbol{p}}(l)\right) \quad (16)$$

where $N_p\setminus\boldsymbol{q}$ means all the pixels at the neighbourhood of pixel $\boldsymbol{p}$ except for $\boldsymbol{q}$, to which the message is being sent. It should be noted that for a pixel $\boldsymbol{q}$ should only receive messages about disparities that belong to its candidate disparity set, i.e. $d\in D_q$.

The details of the algorithm are avoided here and readers are referred to Sun et al. (2003) and (2005) for more details. The message sending iterates several times until all the nodes receive the complete messages from the other nodes. Then, at the end, the belief at each node $\boldsymbol{p}$ about any disparity candidate d (i.e. $B_p(d)$) can be computed as follows.

$$B_p(d)=\gamma(\boldsymbol{p},d)\prod_{\boldsymbol{q}\in N_p}msg_{\boldsymbol{q}\to\boldsymbol{p}}(d) \quad (17)$$

Therefore, the disparity candidate which maximizes the belief is the optimal one ($\hat{d}_p$), which together with the optimal disparities at other pixels maximizes the posterior of Equation (13), or equivalently, minimizes the energy of Equation (7).

$$\hat{d}_p=\arg\max_{d\in D_p}\left(B_p(d)\right) \quad (18)$$

The matching algorithm is repeated once considering the left image as the reference and once in reverse way. At the end, left-right cross-checking is also performed to remove the outliers.

5. EXPERIMENTAL RESULTS

In order to test different aspects of the proposed dense matching algorithm, some training stereo images belonging to the Middlebury Stereo benchmark were used (Scharstein and Szeliski, 2002; Scharstein and Szeliski, 2003; Scharstein and Pal, 2007; Hirschmüller and Scharstein, 2007). In future, more tests will be performed on the 2014 datasets of the benchmark, which will allow the comparison of this algorithm with other state-of-the-art ones.

For hypothesis generation, orientation threshold of T=15 degrees was used (in Equation (3)), and the maximum arc-length in the curve was considered to decide the arc-length threshold σ in Equation (4). The experiments on the stereo images of Figure 6 showed that the ground-truth disparity value of more than 86±8% of the pixels existed among the candidate disparities hypothesized for those pixels. The reason of failure at other pixels was the fact that no hypothesis was generated at all for those pixels. They are, mainly, the pixels located at texture-less or uni-color areas of the image. At these areas, the orientation at the intrinsic curve is undefined since there is neither intensity nor gradient-of-intensity changes. Therefore, these points would be left with no candidate hypothesis, unless some were found at the hypothesis extension step. That is, a pixel is locally located in the texture-less area but its close surrounding area has texture variations.

In addition, the size of the set of candidate disparities for each pixel was, in average, 17±3% of the size of the full disparity search range. Since the computational complexity of most matching algorithms is linearly dependant to the size of the disparity search space, such reduction of the search space can be proportional to the increase of matching efficiency. In case of belief propagation, the complexity of optimization is linearly dependant on the size of disparity search space. That is our algorithm increased the speed of LBP by 83% compared to the case when the whole disparity search space must have been considered.

Regarding the proposed method to detect potential occlusions, a test was performed on the stereo images. For each pair, the function Θ, as defined in Equation (6), was calculated at non-occluded ground-truth pixels. Figure 7 shows the value of this function for the image of dolls in Figure 6 (note that the Θ values were ordered ascending). The same results were obtained for other images. In average for 92±1% of the non-occluded pixels, the value of Θ was exactly zero. Therefore, the consideration of this function to estimate the occlusion term of the energy function (E_{occ}) as a soft prior (not a hard/binary constraint) was logical.

Figure 8 shows the computed disparity maps using the proposed method. In the experiments of this study, disparity refinement was not performed, e.g. median filtering or sub-pixel estimation and gap-filling using interpolation techniques; because the objective of these experiments was to check the accuracy achievable only by the matching algorithm. Under each disparity map at Figure 8, the accuracy of matching is denoted with two values R_1 and R_2. R_1 is the percentage of pixels whose computed disparity values differ from their respective ground-truth disparities less than 1 pixel. R_2 is the percentage of pixels whose computed disparity values have more than 2 pixels difference from their respective ground-truth disparities. In average, the matching algorithm succeeded to match 92.8% of

pixels with less than 1 pixel difference from the ground-truth. It should be noted that the estimated disparity maps were approximately from 4% to 20% less dense than the ground-truth maps mainly because of the pixels for which no candidate disparity was hypothesized.

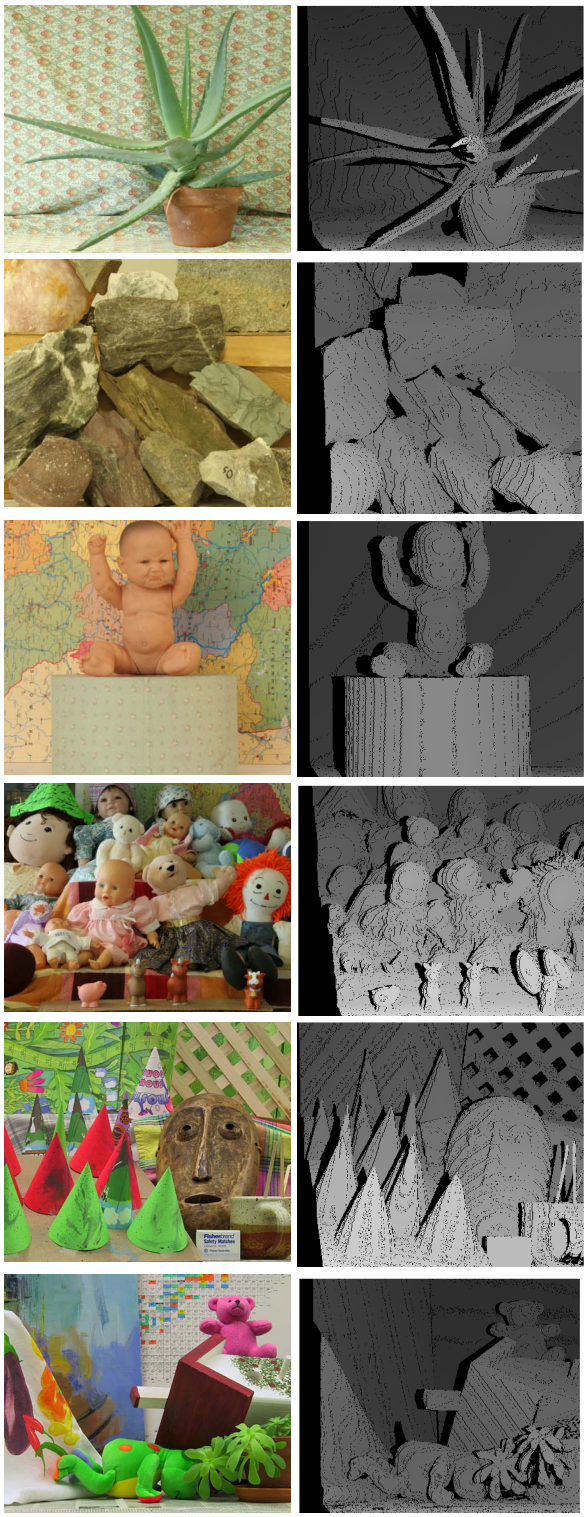

Figure 6. Test stereo datasets: the left images and their non-occluded ground-truth disparity maps to which left-right cross checking was applied.

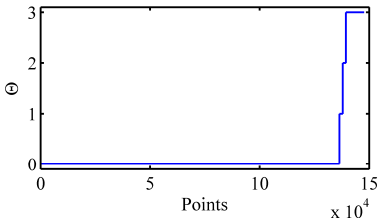

Figure 7. Values of curvature similarity Θ at non-occluded pixels of a stereo pair

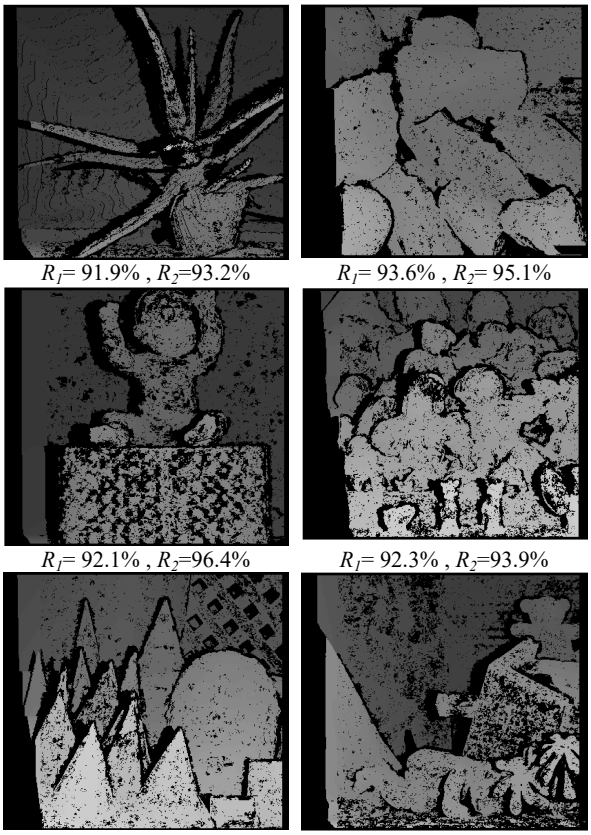

Figure 8. Estimated disparity maps using the proposed matching technique and their accuracies.

6. FUTURE WORK AND CONCLUSION

In this study, we revisited the concepts of intrinsic curves and studied their characteristics in order to propose an efficient dense stereo matching algorithm. The first objective of this study was to reduce the disparity search range with a simple search at the space of the intrinsic curves and without requiring computationally complex techniques such as hierarchical matching. The suggested hypothesis generation technique considered both spatial and photometric characteristics of pixels to propose the candidate disparities. The second objective was to reduce the ambiguities due to occluded pixels of images by integrating the occlusion-detection clues explicitly into the global energy function as a soft prior. These clues were also derived using the concepts of intrinsic curves and their capacity to manifest occlusions.

In the future, modifications will be applied to the hypothesis generation technique so that the orientation thresholds can be decided adaptively based on image observations, e.g. intensity changes. This modification will help to fill the gaps at non-occluded pixels, and, at the same time, to keep generating efficient disparity candidate sets with relatively small size. Furthermore, the details of the matching algorithm using LBP will be investigated. For instance, different matching cost functions will be tested to choose the one which best fits to this framework. In addition, the results by different algorithms of LBP such as sum-product will be evaluated and compared. In the future, when the matching algorithm will be tested comprehensively, the results obtained from the datasets of the Middlebury Stereo benchmark and ISPRS Benchmark on High Density Aerial Image Matching will be submitted for further evaluation.

ACKNOWLEDGEMENTS

This study is supported in part by grants from: Centre de Géomatique du Québec, Fonds de Recherche Québécois sur la Nature et les Technologies, and Natural Sciences and Engineering Research Council of Canada.

REFERENCES

Bleyer, M., Rhemann, C. and Rother, C., 2011. PatchMatch stereo-stereo matching with slanted support windows. In *Proceedings of the British Machine Vision Conference*, pp. 1-11.

Boykov, Y., Veksler, O. and Zabih, R., 1998. Markov random fields with efficient approximations. In: Proceedings of IEEE Conference on Computer vision and pattern recognition, pp. 648-655.

Boykov, Y., Veksler, O. and Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE T Pattern Anal*, 23(11), Edinburgh, pp.1222-1239.

Cox, I.J., Hingorani, S.L., Rao, S.B. and Maggs, B.M., 1996. A maximum likelihood stereo algorithm. *Comput Vis Image Und*, 63(3), pp.542-567.

Hirschmüller, H., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE T Pattern Anal*, 30(2), pp.328-341.

Hirschmüller, H. and Scharstein, D., 2007. Evaluation of cost functions for stereo matching. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, pp. 1-8.

Hosni, A., Bleyer, M., Gelautz, M. and Rhemann, C., 2009. Local stereo matching using geodesic support weights. In: *Proceedings of IEEE International Conference on Image Processing*, Cairo, pp. 2093-2096.

Hosni, A., Rhemann, C., Bleyer, M., Rother, C. and Gelautz, M., 2013. Fast cost-volume filtering for visual correspondence and beyond. *IEEE T Pattern Anal*, 35(2), pp. 504-511.

Geiger, A., Roser, M. and Urtasun, R., 2011. Efficient large-scale stereo matching. *Lecture Notes in Computer Science*, 6492, pp. 25-38.

Gurbuz, Y.Z., Alatan, A.A. and Cigla, C., 2015. Sparse recursive cost aggregation towards O(1) complexity local stereo matching. In: *Proceedings of IEEE Conference on Signal Processing and Communications Applications*, Malatya, pp. 2290-2293.

Ohta, Y. and Kanade, T., 1985. Stereo by intra-and inter-scanline search using dynamic programming. *IEEE T Pattern Anal*, 2, pp.139-154.

Mozerov, M.G. and van de Weijer, J., 2015. Accurate stereo matching by two-step energy minimization. *IEEE T Image Process*, 24(3), pp.1153-1163.

Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nesic, N., Wang, X. and P. Westling, 2014. High-resolution stereo datasets with subpixel-accurate ground truth. In: *Proceedings of German Conference on Pattern Recognition*, Münster, pp. 1-12.

Scharstein, D. and Pal, C., 2007. Learning conditional random fields for stereo. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, pp. 1-8.

Scharstein, D. and Szeliski, R., 2003. High-accuracy stereo depth maps using structured light. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Madison, WI, pp.195-202.

Sinha, S.N., Scharstein, D. and Szeliski, R., 2014. Efficient high-resolution stereo matching using local plane sweeps. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, pp. 1582-1589.

Sun, J., Zheng, N.N. and Shum, H.Y., 2003. Stereo matching using belief propagation. *IEEE T Pattern Anal*, 25(7), pp.787-800.

Sun, J., Li, Y., Kang, S.B. and Shum, H.Y., 2005. Symmetric stereo matching for occlusion handling. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, pp. 399-406.

Tomasi, C. and Manduchi, R., 1998. Stereo matching as a nearest-neighbor problem. *IEEE T Pattern Anal*, 20(3), pp. 333-340.

Yoon, K.J. and Kweon, I.-S., 2006. Adaptive Support-Weight Approach for Correspondence Search, *IEEE T Pattern Anal*, 28(4), pp. 650-656.

Veksler, O., 2006. Reducing search space for stereo correspondence with graph cuts. In *Proceedings of the British Machine Vision Conference*, pp. 73.1-73.10.

Zabih, R. and Woodfill, J., 1994. Non-parametric local transforms for computing visual correspondence. *Lecture Notes in Computer Science*, 801, pp. 151–158.

Zitnick, C.L. and Kanade, T., 2000. A cooperative algorithm for stereo matching and occlusion detection. *IEEE T Pattern Anal*, 22(7), pp.675-684.