Visual Correspondence Using Energy Minimization and Mutual Information

Junhwan Kim Vladimir Kolmogorov Ramin Zabih
Computer Science Department
Cornell University
Ithaca, NY 14853

Abstract

We address visual correspondence problems without assuming that scene points have similar intensities in different views.This situation is common, usually due to non-lambertian scenes or to differences between cameras. We use maximization of mutual information, a powerful technique for registering images that requires no a priori *model of the relationship between scene intensities in different views. However, it has proven difficult to use mutual information to compute dense visual correspondence. Comparing fixed-size windows via mutual information suffers from the well-known problems of fixed windows, namely poor performance at discontinuities and in low-texture regions. In this paper, we show how to compute visual correspondence using mutual information without suffering from these problems. Using a simple approximation, mutual information can be incorporated into the standard energy minimization framework used in early vision. The energy can then be efficiently minimized using graph cuts, which preserve discontinuities and handle low-texture regions. The resulting algorithm combines the accurate disparity maps that come from graph cuts with the tolerance for intensity changes that comes from mutual information.*

1. Introduction

The visual correspondence problem is to compute the pairs of pixels from two images that result from the same scene element. Since the correspondence problem is inherently ill-posed, assumptions must be made regarding scene reflectance and structure. It is common to assume that a given scene element will result in similar intensities in different views (the "constant brightness assumption"). However, this holds only when the surfaces in the scene are lambertian and the mapping from reflectance to intensity captured by the camera (e.g. camera gain and bias) are identical among different views. When the constant brightness assumption is violated, for example in the presence of non-lambertian reflectance or different camera gains or biases, corresponding scene elements in different images can be poorly correlated, leading to incorrect results. For example, Figure 1 shows an image pair that has been synthetically altered, by giving the right image a negative gain. A traditional stereo algorithm gives very poor results.

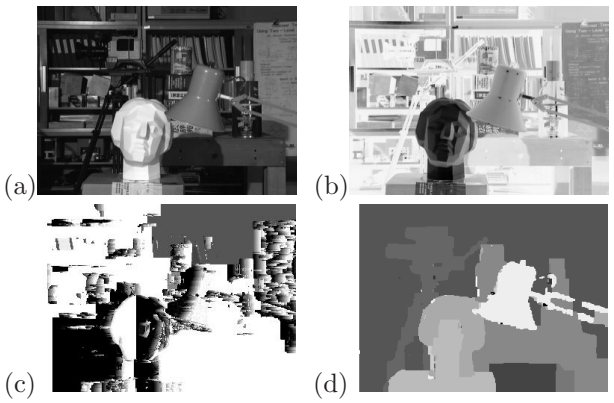

Figure 1: Comparison of our algorithm and a standard stereo algorithm: (a) Left image (b) Right image, synthetically altered (c) Result from a traditional stereo algorithm [3] (d) Result from our method

The correspondence problem can be formally defined as follows. Let $\mathcal{P}$ denote the set of pixels in the primary image and let $I_1 = \{I_1(p) \mid p \in \mathcal{P}\}$ and $I_2 = \{I_2(p) \mid p \in \mathcal{P}\}$ be the intensities in the primary and secondary images.[1]The quantity to be estimated is the disparity configuration $f = \{f_p \mid p \in \mathcal{P}\}$ on the primary image. Each f_p represents the correspondence between the pixel p in the primary image and the pixel $p + f_p$ in the secondary image, i.e.,

[1]There is no fundamental asymmetry in the correspondence problem, but most algorithms treat the input images asymmetrically. We will use an asymmetric treatment throughout this paper, then describe in the final section how to extend our work to a symmetric treatment.

the pixel whose 2D coordinates are obtained by adding the disparity f_p to the 2D coordinates of p. The constant brightness assumption is $I_1(p) \simeq I_2(p + f_p)$. In this paper, we address the correspondence problem where $I_1(p) \simeq F(I_2(p + f_p))$, or more generally, $F(I_1(p), I_2(p + f_p)) \simeq 0$, where F is an *a priori* unknown function. For instance, $I_1(p) = \alpha I_2(p + f_p)) + \beta$ implies different gain and bias for the two cameras.

There are two broad classes of algorithms for computing visual correspondence (see [14] for a recent survey for stereo). Local algorithms estimate correspondence independently at each pixel, typically using correlation with fixed size windows. Global algorithms find the best disparity configuration f. In general, while local algorithms are faster, the global algorithms give the best results [14]. The difference is most striking at disparity discontinuities and in low-texture regions. Most global algorithms are based on a standard energy minimization framework, which is closely related to Markov Random Fields (MRF's) [5, 10].

1.1. Energy minimization

Energy minimization algorithms define the best disparity configuration f to be the one that minimizes the energy, which consists of a smoothness term and a data term

$$E(f) \quad = \quad E_{\text{smooth}}(f) + E_{\text{data}}(f). \qquad (1)$$

The smoothness term imposes a penalty for configurations that violate spatial smoothness. The data term imposes a penalty for configurations that are inconsistent with the observed data I_1, I_2. The data term is where the appearance of corresponding scene elements is used, and will be the focus of our work. The standard data term used is

$$E_{\text{data}}(f) = \sum_p D_p(f_p), \qquad (2)$$

where D_p is a measure of pixel dissimilarity between $I_1(p)$ and $I_2(p + f_p)$. Nearly all work on energy minimization makes the constant brightness assumption, and has $D_p(f_p) = \rho(I_1(p), I_2(p + f_p))$, where ρ is some distance measure. We will describe a few common choices of ρ when we survey related work in Section 2.

The most difficult problem in energy minimization is its computational cost, since it involves a highly non-convex function in a search space with many thousands of dimensions. The energy can be efficiently minimized using graph cuts [3, 8, 12] as long as it is of a certain form. The data term must be of the form given in Eq (2), i.e. a sum over pixels; in addition, there are some restrictions on the smoothness term. Energy minimization methods based on graph cuts generate some of the best results for visual correspondence [14], despite making the constant brightness assumption.

1.2. Mutual information (MI)

Our approach to the correspondence problem relies on maximization of mutual information. Mutual information (MI) was invented by Shannon [15], and popularized in computer vision by Viola and Wells [17]. It has been primarily used for registration problems, where the goal is to find the transformation that makes one image most similar to another. Mutual information is used as a similarity measure between images, and the transformation that maximizes the mutual information is found by some variant of gradient descent [17]. The key advantage of mutual information is its ability to easily handle complex relationships between the intensities in the two images. It requires no *a priori* model of the relationship between scene intensities in different views, and thus can even register medical images from different modalities (such as CT and MR) [19].

Mutual information is therefore a very natural technique to use for visual correspondence. However, this has proven difficult to accomplish. It is easy to incorporate mutual information into a local algorithm with fixed size windows. This has been done by Egnal [4] (see Section 2 for more discussion). However, this suffers from all the standard problems of fixed window methods, and gives poor results.

1.3. Overview

In this paper we show how to incorporate mutual information into an energy minimization algorithm for computing visual correspondence. Our key technical contribution is to develop a data term that uses mutual information, while ensuring that the resulting energy function can be efficiently minimized. This is non-trivial, because mutual information is not naturally defined as a sum over pixels, which is required for efficient minimization. However, using a Taylor series approximation we can rewrite the mutual information as a sum over pixels, and then use graph cuts to efficiently maximize it. This allows us to combine the accurate disparity maps that come from graph cuts with the tolerance for intensity changes that comes from mutual information.

We begin with a short summary of some related work. In Section 3, we discuss mutual information as a data term and simplify it slightly. In Section 4, we show how to rewrite this data term as a sum over pixels of a certain D_p^{MI}. In Section 5, we demonstrate that D_p^{MI} in fact generalizes a standard data term from the well-known framework of maximum *a posteriori* estimation

of MRF's [5, 10]. Preliminary experimental results are given in Section 7. For real data with ground truth, the performance of our method is roughly comparable to several other energy minimization methods cited in [14]. However, artificially distorting the intensities in these images has only a small effect on our method, while it has a dramatic effect on previous methods.

2. Related work

2.1. Stereo matching costs

The visual correspondence problem has been extensively studied in the computer vision community. Relaxing the constant brightness assumption has been most heavily studied for stereo, as it is common to have cameras with different gain and bias. Since our work is novel primarily in terms of the matching cost, we will focus on related work that addresses this. Readers are referred to [14] for a survey and taxonomy of stereo.

The most common matching costs include the sum of L^1 differences [16], L^2 differences [11], or truncated L^2 differences [1]. These costs are sensitive to camera gain and bias. It is also possible to first compute a local quantity that is insensitive to gain and bias and then perform correlation. This has been done using ordering information [21] or properties of the intensity gradient [13]. Another approach [6] eliminates photometric effects using a spatial coherence multiplier in the matching process.

Our work is distinctive because of its generality. Mutual information allows for a wide range of relationships between intensities from corresponding scene elements. It merely assumes that there is a consistent relationship between them, which we do not or cannot specify *a priori*.

2.2. Mutual information

Mutual information has been popularized in computer vision by Viola and Wells [17]. It can be used for for pose estimation, object recognition, shape from shading, and lightness compensation. However, its primary use is for registration, typically using affine transforms. [18] and [20], among others, use maximization of mutual information for nonrigid registration. They represent nonrigid registration using thin-plate splines, which do not preserve discontinuity. They evaluate the disparities only for sparsely sampled pixels and approximate the disparities for in-between pixels using the radial basis function. In contrast, we evaluate disparities for every pixel, thus preserving discontinuities.

It is, of course, possible to use any registration technique to compute visual correspondence, simply by applying it to fixed windows centered at each pixel. Egnal [4] used mutual information in this way. This approach suffers from the standard limitations of fixed window methods, namely poor performance at discontinuities, and in low-texture regions. These problems can be overcome using energy minimization.

3. MI as a data term

It is natural to use mutual information (MI) as a data term in the energy function,

$$E_{\text{data}}^{\text{MI}}(f) = -\text{MI}(I_1, I_2, f). \tag{3}$$

(This is negative because mutual information is maximized, while energy must be minimized.) Here $\text{MI}(I_1, I_2, f)$ is the mutual information between the two images I_1 and I_2 given the disparity f. We can express $\text{MI}(I_1, I_2, f)$ as the sum of the entropy of I_1 and the entropy of I_2 minus the joint entropy of I_1 and I_2,

$$\begin{aligned}
&\text{MI}(I_1, I_2, f) = h(I_1) + h(I_2, f) - h(I_1, I_2, f) \\
&= -\int_0^1 di P_{I_1}(i) \log P_{I_1}(i) - \int_0^1 di P_{I_2,f}(i) \log P_{I_2,f}(i) \\
&\quad + \int_0^1 \int_0^1 di_1 di_2 P_{I_1,I_2,f}(i_1, i_2) \log P_{I_1,I_2,f}(i_1, i_2), \quad (4)
\end{aligned}$$

where we define

$$P_{I_1}(i) = \frac{1}{|\mathcal{P}|} \sum_p g_\psi(i - I_1(p)).$$

Following [17] we used Parzen estimation with a Gaussian distribution; $g_\psi(x - \mu)$ denotes a Gaussian distribution with mean μ and variance ψ. Similarly, we define

$$\begin{aligned}
P_{I_2,f}(i) &= \frac{1}{|\mathcal{P}|} \sum_p g_\psi(i - I_2(p + f_p)), \\
P_{I_1,I_2,f}(i_1, i_2) &= \frac{1}{|\mathcal{P}|} \sum_p g_\psi((i_1, i_2) - (I_1(p), I_2(p + f_p))).
\end{aligned}$$

Here, $g_\psi(x - \mu)$ denotes an $n(= 2)$ dimensional Gaussian distribution, with mean μ and covariance matrix ψ. Throughout this paper, we will use a diagonal matrix for ψ.[2]

Note that $h(I_1)$ in Eq (4) does not depend on f. $h(I_2, f)$ also is almost constant; f merely redistributes when calculating $P_{I_2,f}(i)$ except for occlusions or many-to-one matchings from I_1 to I_2. If the I_1 to I_2 matching is one to one, $P_{I_2,f}(i)$ is a constant. We will therefore regard $h(I_2, f)$ as a constant.

[2] This does not mean that the left image and the right image are independent. It means that the *noise* of the left image and that of the right image are independent.

To summarize, ignoring some constants, the mutual information data term has the form

$$E^{\mathrm{MI}}_{\mathrm{data}}(f) = -\iint di_1 di_2 P_f(i_1,i_2)\log(P_f(i_1,i_2)), \quad (5)$$

where we use $P_f(\cdot,\cdot)$ to denote $P_{I_1,I_2,f}(\cdot,\cdot)$ in order to simplify the notation.

4. Approximating MI

The key technical challenge is to convert the mutual information (MI) data term $E^{\mathrm{MI}}_{\mathrm{data}}$ into a sum over pixels as in Eq (2). Once this is done, we can efficiently minimize the energy using graph cuts. We now show how to find a D^{MI}_p such that

$$E^{\mathrm{MI}}_{\mathrm{data}} \simeq \sum_p D^{\mathrm{MI}}_p(f_p).$$

We use the Taylor expansion for $F(x) = x\log x$:

$$\begin{aligned} F(x) &= F(x^0) + F'(x^0)(x-x^0) + O((x-x^0)^2) \\ &= x^0\log x^0 + (1+\log x^0)(x-x^0) + O((x-x^0)^2) \\ &= -x^0 + (1+\log x^0)x + O((x-x^0)^2). \end{aligned}$$

Consider two arbitrary disparity configurations f, f^0 (there is no requirement that they be similar). Using this Taylor expansion on $E^{\mathrm{MI}}_{\mathrm{data}}(f)$ we have

$$\begin{aligned} & -\iint di_1 di_2 \log(P_f(i_1,i_2))P_f(i_1,i_2) \quad (6) \\ \simeq & \iint di_1 di_2 \Big(P_{f^0}(i_1,i_2) - (1+\log(P_{f^0}(i_1,i_2)))P_f(i_1,i_2)\Big) \\ = & \iint di_1 di_2 P_{f^0}(i_1,i_2) - \iint di_1 di_2 P_f(i_1,i_2) \\ & -\iint di_1 di_2 \log(P_{f^0}(i_1,i_2))P_f(i_1,i_2). \end{aligned}$$

Since the first two terms are 1 by definition of probability, this can be rewritten as

$$\begin{aligned} & -\iint di_1 di_2 \log(P_{f^0}(i_1,i_2))P_f(i_1,i_2) \\ = & -\iint di_1 di_2 \Big(\log(P_{f^0}(i_1,i_2)) \\ & \cdot \frac{1}{|\mathcal{P}|}\sum_p g_\psi\left((i_1,i_2) - (I_1(p), I_2(p+f_p))\right)\Big) \\ = & \sum_p -\frac{1}{|\mathcal{P}|}\iint di_1 di_2 \Big(\log(P_{f^0}(i_1,i_2)) \\ & \cdot g_\psi((i_1,i_2) - (I_1(p), I_2(p+f_p)))\Big). \quad (7) \end{aligned}$$

In order for the first order approximation of Eq (6) to be valid, we need to have $|x-x^0|/\min(x,x^0) \ll |1+\log x^0|$, that is, x close to x^0. This condition does not imply that f is close to f^0. x and x^0 are 2D intensity histograms of pixels that are alleged to correspond by a disparity configuration (f or f^0). We allow a large set of pixels to go from one bin to another then, as long as another set of pixels replaces them. Since we are usually dealing with a large number of pixels, it is reasonable to expect this to hold.[3]

Our desired result follows directly:

$$\begin{aligned} D^{\mathrm{MI}}_p(f_p) = & -\frac{1}{|\mathcal{P}|}\iint di_1 di_2 \Big(\log(P_{f^0}(i_1,i_2)) \\ & \cdot g_\psi((i_1,i_2) - (I_1(p), I_2(p+f_p)))\Big). \quad (8) \end{aligned}$$

To summarize, by using the approximation in Eq (6) we can obtain a mutual information data term that is in the standard sum of pixels form. This yields an energy functional that can be efficiently minimized. Note that the data term $D^{\mathrm{MI}}_p(f_p)$ depends on the current disparity map f^0. We will update this term as the algorithm iterates.

5. Reduction to MAP-MRF

In this section we show how our expression for the data term can be justified in the MAP-MRF framework [5] under certain assumptions. Let $P(i_1)$ be the distribution of intensities in the left image and assume that the current disparity (f^0) is the true disparity. Let us also assume that the right image is generated by adding an independent noise described by the distribution $\Pr(i_2|i_1)$ to the left image warped by f^0. Then the joint histogram $P_{f^0}(i_1,i_2)$ will be approximately $P(i_1)\Pr(i_2|i_1)$. For simplicity we will assume that we have enough samples so that the following formula is exact:

$$P_{f^0}(i_1,i_2) = P(i_1)\Pr(i_2|i_1).$$

Now let us compute our data term given the current disparity map f^0. If we use a smoothing kernel with $\psi \to 0$ (this is reasonable since we assume that we have enough samples), Eq (8) reduces to

$$\begin{aligned} D^{\mathrm{MI}}_p(f_p) &= -\frac{1}{|\mathcal{P}|}\log(P_{f^0}(I_1(p), I_2(p+f_p))) \\ &= -\frac{1}{|\mathcal{P}|}\left(\log(P(i_1)) + \log(\Pr(I_2(p+f_p)|I_1(p)))\right). \end{aligned}$$

The first term can be omitted since it does not depend on f_p, resulting in

$$D^{\mathrm{MI}}_p(f_p) = -\frac{1}{|\mathcal{P}|}\log(\Pr(I_2(p+f_p)|I_1(p))).$$

[3] The difference between the left and right sides of the approximation in Eq (6) is that f in $\log(P_f(i_1,i_2))$ is replaced by f^0. This is reminiscent of the difference between the implicit and explicit methods in finite-difference methods [7].

This is a classical expression used in the MAP-MRF framework (except for the constant $-\frac{1}{|\mathcal{P}|}$). Thus our expression for the data term coincides with the MAP-MRF expression. In particular, if the noise $\Pr(i_2|i_1)$ is Gaussian (i.e. $\Pr(i_2|i_1) \sim \exp(-(i_2 - i_1)^2/2\sigma^2)$), then Eq (8) reduces to

$$D_p^{\mathrm{MI}}(f_p) = \mathrm{const} \cdot (I_1(p) - I_2(p + f_p))^2.$$

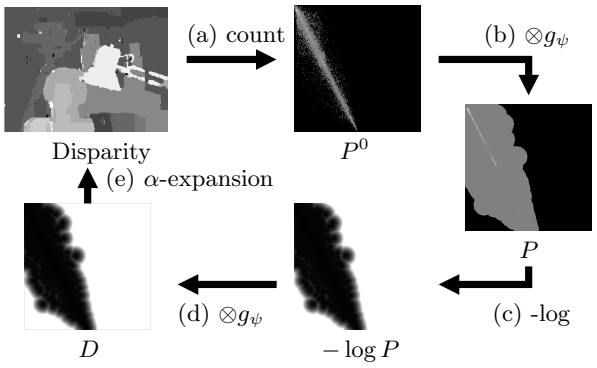

Figure 2: An example that depicts the construction of the MI data term. Dark black in P and P^0 represents 0. See section 4 for detailed explanation.

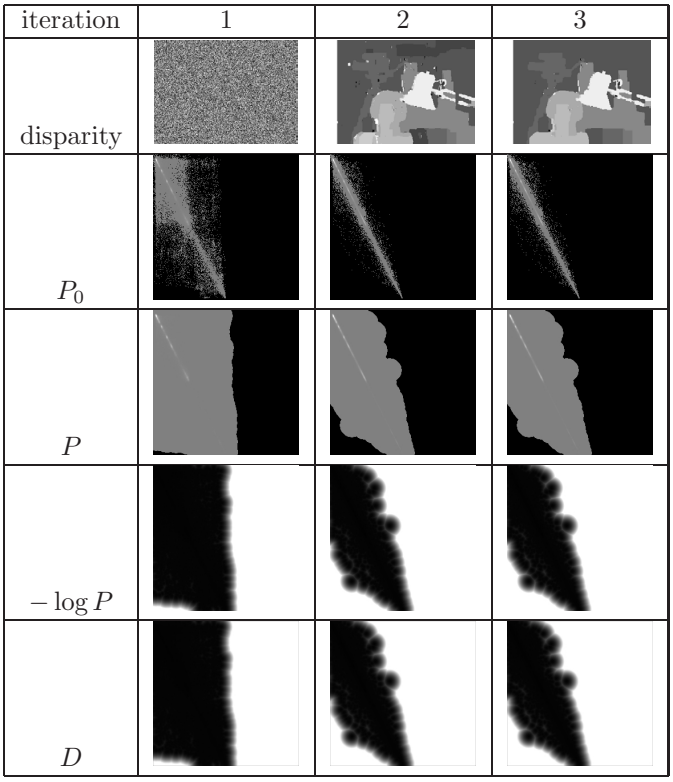

Figure 3: An example of key data values over three iterations of our algorithm.

6. Our algorithm

Now that we have obtained a data term of the correct form, it is straightforward to compute correspondence. All we need is to specify E_{smooth}, and to select an energy minimization algorithm. For $E_{\mathrm{smooth}}(f)$, we use the Potts model energy function, $\sum_{p,q \in N} V_{p,q}(f_p, f_q)$, where $V_{p,q}(f_p, f_q) = u_{\{p,q\}} \cdot T[f_p \neq f_q]$. Here $T[\cdot]$ is 1 if its argument is true and 0 otherwise. The $u_{\{p,q\}}$ multiplier can be interpreted as the cost of a discontinuity between p and q [3]. We use the α-expansion algorithm for energy minimization [3]. Algorithm 1 depicts α-expansion with the data term given by Eq (8). It iterates between (a) constructing the data term D_p^{MI} from the current f (Line: 4) and (b) finding a new f given probability and data term (Lines: 5-14).

The construction of the data term D_p^{MI} merits a more detailed description. Since $P_{f^0}(i_1, i_2)$ doesn't depend on p or f_p, we should compute it only once. We first compute the histogram of allegedly corresponding pixels, or $P^0_{f^0}(i_1, i_2)$, for $\phi = \lim_{\sigma \to 0} \mathrm{diag}(\sigma, \sigma)$ by simple counting (see Figure 2(a));

$$P^0_{f^0}(i_1, i_2) = \frac{1}{|\mathcal{P}|} \sum_p T[(i_1, i_2) = (I_1(p), I_2(p + f_p^0))]. \quad (9)$$

Then we apply Gaussian convolution to obtain the probability distribution, $P_{f^0}(i_1, i_2)$ (see Figure 2(b));

$$P_{f^0}(i_1, i_2) = P^0_{f^0}(i_1, i_2) \otimes g_\psi(i_1, i_2). \quad (10)$$

In other words, Gaussian convolution computes the distribution from the samples.

If we have the correct disparity, and there are no differences in gain and bias, this probability distribution would be a 45 degree line through the origin. We discretize i_1 and i_2 into N_{int} values and compute $P_{f^0}(i_1, i_2)$ for each i_1 and i_2 in $O(|\mathcal{P}|)$ for Eq (9) and $O(N_{\mathrm{int}}^2 \log N_{\mathrm{int}})$ for Eq (10) using the FFT and taking advantage of the fact that Gaussian convolution is linearly separable.

We apply Gaussian convolution once more, this time to calculate the data term $D_p^{\mathrm{MI}}(f_p)$ (see Figure 2(d));

$$D(i_1, i_2) = -\frac{1}{|\mathcal{P}|} \log(P_{f^0}(i_1, i_2)) \otimes g_\psi(i_1, i_2), \quad (11)$$

$$D_p^{\mathrm{MI}}(f_p) = D(I_1(p), I_2(p + f_p)). \quad (12)$$

The overall time complexity is $O(N_{\mathrm{int}}^2 \log N_{\mathrm{int}})$ for the convolution and $O(|\mathcal{P}||\mathcal{L}|)$ for Eq (12), where $\mathcal{L}$ is the

set of all possible labels. Throughout this paper, we use $N_{\text{int}} = 256$, unless otherwise indicated.

In the final step we use the α-expansion algorithm with this data term to compute a new disparity (see Figure 2(e)). We iterate this cycle until convergence. See Figure 3 for an example.

Note that it is possible to update D_p^{MI} more frequently, since it depends upon the current disparity configuration f. However, the results we have obtained experimentally do not improve substantially with more frequent updates.

Algorithm 1 α-expansion with our mutual information data term

Require: I_1, I_2
Ensure: $f^0 = \text{argmin } E(f)$

```
Start with an arbitrary configuration f^0
repeat
    Set SuccessOut := 0
    ∀p calculate f_p, D_p^MI(f_p) from I_1, I_2, f^0
    repeat
        Set SuccessIn := 0
        for each label α ∈ L do
            Find f̂ = argmin E(f') among f' within one
            α-expansion of f^0
            if E(f̂) < E(f^0) then
                Set f^0 := f̂
                Set SuccessIn := 1 and SuccessOut := 1
            end if
        end for
    until SuccessIn = 1
until SuccessOut = 1
```

7. Experiments

To verify that our algorithm is insensitive to the relationship between left and right images, we apply various transforms to the intensities of one of the images and run our algorithm (see Figure 5). As expected our algorithm gives near-identical results when the images are transformed, unlike the traditional matching cost (see Figure 1(c)). We also tested the performance on stereo images in the presence of specularity, which is the most dramatic form of non-lambertian reflectance, and the most serious violation of the constant brightness assumption. Figure 4 shows some promising preliminary results.

It is also instructive to look at the performance of our algorithm on the real images with ground truth described in [14]. The statistics are shown in Figure 5.The running times for most of our test cases does not exceed 2-3 minutes on a Pentium IV processor. Our method converges rather quickly within a few iterations (see Figure 3). This is partly because even if the current disparity map is not precisely correct, the kernel estimate of probability is not terribly wrong when the images are sparsely textured. On richly textured images, it is possible that better results could be obtained by a nonconvex method where the smoothing of the kernel estimate is gradually reduced as the disparities become more correct. However, note that we obtain good results on the richly textured SRI tree sequence shown in the bottom row of figure 5.

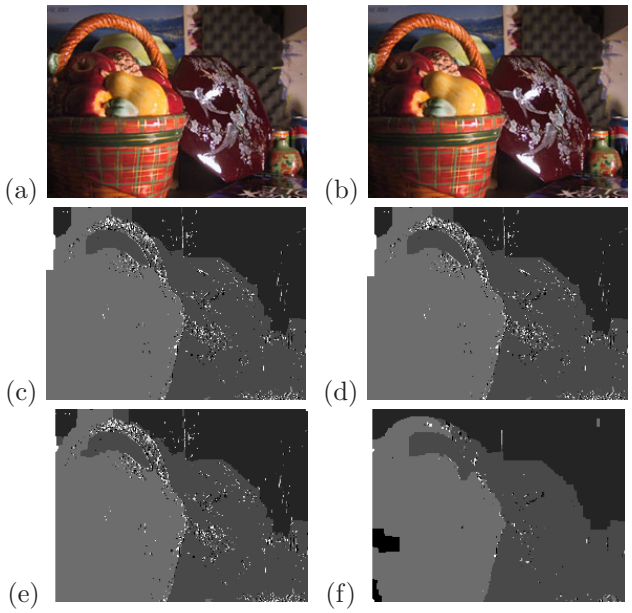

Figure 4: Our results on an image pair with specularity. (a) Left image (b) Right image (c) α-expansion with truncated L^2 difference data term (d) α-expansion with L^2 difference data term (e) α-expansion with L^1 difference data term (f) α-expansion with mutual information data term

Overall, our algorithm produces results that are comparable to several other energy minimization approaches, and are significantly better than standard correlation-based methods. However, unlike previous methods, our algorithm is stable under a very broad range of intensity transformations.

8. Extensions

Both mutual information and the correspondence problem are, by definition, symmetric. Despite this our formulation so far treats left image and right image asymmetrically. Nevertheless, a framework due to Kolmogorov and Zabih [9] allows us to extend our formulation to treat the images symmetrically, and also to properly treat occlusions. Following their notation, let

$\mathcal{A}$ be the set of pairs of pixels that may potentially correspond. Let $A(f)$ be the set of active assignments according to the configuration f. The energy for a configuration f is given by

$$E(f) = E_{\text{data}}(f) + E_{\text{occ}}(f) + E_{\text{smooth}}(f), \qquad (13)$$

where $E_{\text{data}}(f) = \sum_{\langle p,q\rangle \in A(f)} D(\langle p,q\rangle)$ and $D(\langle p,q\rangle) = (I_1(p) - I_2(q))^2$. Following similar procedures to those used in section 3, we can derive a mutual information data term

$$E^{\text{MI}}_{\text{data}}(f) \simeq \sum_{\langle p,q\rangle \in A(f)} D^{\text{MI}}(\langle p,q\rangle), \qquad (14)$$

$$D^{\text{MI}}(\langle p,q\rangle) = -\frac{1}{|A(f)|}\iint di_1 di_2 \Big(\log(P_{f^0}(i_1,i_2)) \cdot g_\psi((i_1,i_2) - (I_1(p), I_2(q)))\Big), \qquad (15)$$

where

$$P_{f^0}(i_1,i_2) = \frac{1}{|A(f^0)|} \sum_{\langle p,q\rangle \in A(f^0)} g_\psi((i_1,i_2) - (I_1(p), I_2(q))).$$

Notice that we modify only $D(\langle p,q\rangle)$ with all other terms unchanged.

Acknowledgments

We thank Amy Gale for proofreading and the anonymous reviewers for their constructive critiques. We also thank Rahul Swaminathan *et al.* for providing us the input data in figure 4, and Daniel Scharstein and Rick Szeliski for providing us with some of the imagery and the ground truth images in figure 5. This work was supported by NSF grants IIS-9900115 and CCR-0113371 and by a grant from Microsoft Research.

References

[1] M.J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow-fields. *Computer Vision and Image Understanding*, 63(1):75–104, January 1996.

[2] Yuri Boykov and Vladimir Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in computer vision. In *Proc. EMMCVPR*, volume 2134 of *Lecture Notes in Computer Science*, pages 359–374. Springer-Verlag, September 2001.

[3] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, November 2001.

[4] Geoffrey Egnal. Mutual information as a stereo correspondence measure. Technical Report MS-CIS-00-20, University of Pennsylvania, 2000.

[5] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[6] Michael Gennert. Brightness-based stereo matching. In *International Conference on Computer Vision*, pages 139–143, 1988.

[7] M. Heath. *Scientific Computing: An Introductory Survey.* McGraw Hill, New York, 2002.

[8] H. Ishikawa and D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. In *European Conference on Computer Vision*, pages 232–248, 1998.

[9] Vladimir Kolmogorov and Ramin Zabih. Visual correspondence with occlusions using graph cuts. In *International Conference on Computer Vision*, pages 508–515, 2001.

[10] S. Li. *Markov Random Field Modeling in Computer Vision.* Springer-Verlag, 1995.

[11] L.H. Matthies, R. Szeliski, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3(3):209–238, September 1989.

[12] S. Roy and I. Cox. A maximum-flow formulation of the n-camera stereo correspondence problem. In *International Conference on Computer Vision*, 1998.

[13] Daniel Scharstein. Matching images by comparing their gradient fields. In *International Conference on Pattern Recognition*, pages 572–575, 1994.

[14] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, April 2002.

[15] Claude E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423, 1948.

[16] E.P. Simoncelli, E.H. Adelson, and D.J. Heeger. Probability distributions of optical flow. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 310–315, 1991.

[17] P.A. Viola and W.M. Wells, III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, September 1997.

[18] Simon K. Warfield *et al.* Intraoperative segmentation and nonrigid registration for image guided therapy. In *MICCAI*, pages 176–185, 2000.

[19] W.M. Wells, III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, March 1996.

[20] M.A. Wirth, J. Narhan, and D Gray. A model for nonrigid mammogram registration using mutual information. In *International Workshop on Digital Mammography*, 2002.

[21] Ramin Zabih and John Woodfill. Non-parametric local transforms for computing visual correspondence. In *European Conference on Computer Vision*, pages 151–158, 1994. Revised version available from www.cs.cornell.edu/rdz.

	Original input	Different camera gain/bias	Intensity correspondence not 1–1	Intensity correspondence not stationary	Ground truth
Left					
	I_1^0	$I_1 = I_1^0/2$	$I_1 = I_1^0$	$I_1 = I_1^0$	
Right					
	I_2^0	$I_2 = I_2^0$	$I_2 = 4(1 - I_2^0/2)^2$	$I_2 = \begin{cases} 4(1 - I_2^0/2)^2 \\ (I_2^0 + 1)/2 \end{cases}$	
Tsukuba	6.39	6.36	6.31	8.36	
Venus	2.37	2.73	4.78	3.40	
Sawtooth	3.63	3.48	5.21	4.65	
Poster	3.53	3.70	3.23	4.05	
Tree					Not available

Figure 5: Our results on synthetically transformed real images, with $\lambda = 0.003, \sigma = 0.0025$. Numbers below the disparities indicate percentage of pixels whose disparities differ from the ground truth by more than 1.