INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Robust Location based Partial Correlation

Zhong-Dan LAN and Roger MOHR

N` 3186

June 5, 1997

THÈME 3

Rapport de recherche

ISSN 0249-6399

Robust Location based Partial Correlation

Zhong-Dan LAN and Roger MOHR

Thème 3 — Interaction homme-machine,
images, données, connaissances

Projet MOVI

Rapport de recherche n° 3186 — June 5, 1997 — 19 pages

Abstract: The visual correspondence problem is a major issue in computer vision. Correlation is a common tool for this problem. Most classical correlation methods fail near the disparity discontinuities, which occur at the boundaries of objects. In this paper, a *partial correlation* technique is proposed to solve this problem. Robust location methods are used to perform this partial correlation. Comparisons are made with other techniques and experimental results validate the approach.

Key-words: vision, matching, correlation, robustness, location, partial occlusion

(Résumé : tsvp)

This work was supported by CNRS, INRIA, INPG and UJF

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN (France)
Téléphone : (33) 04 76 61 52 00 – Télécopie : (33) 04 76 61 52 52

Corrélation partielle basée sur la localisation robuste

Résumé : L'appariement entre deux images est une clé de nombreux problèmes de perception, en particulier celui de la perception stéréoscopique. La corrélation est la solution la plus communément adoptée. La corrélation classique comme ZNCC a des difficultées près des discontinuités qui se présentent à la limite des objets occultants. Dans ce rapport, nous proposons une technique baptisée *corrélation partielle* pour résoudre ce problème. Nous utilisons la *localisation robuste* pour trouver la *bonne* partie des fenêtres à corréler. Nous avons comparé notre méthode avec les méthodes standards et les méthodes non paramétriques. Des expérimentations valident la méthode.

Mots-clé : vision, appariement, corrélation, robustesse, localisation, occultation partielle

Contents

1 Introduction

The visual correspondence problem is a major task in computer vision [11, 2, 6], as it underlies motion and stereo matching. Given two images of the same scene, a pixel in one image corresponds to a pixel in the other if they are projections of the same physical scene element. If they are temporally consecutive, the correspondence determines motion. If they are spatially separated but simultaneous, the correspondence determines stereo disparity.

The use of correlation as a similarity measure between two signals is well known. It is commonly used in stereo vision for the visual correspondence problem and extensive comparisons have been made evaluating different correlation criteria [1].

As indicated in [5, 7], occlusion plays an important rule in stereo matching. In [7], Intille uses Ground Control Points (GCP) to construct the best disparity path. This method gives the best results if GCP are very close to the occlusion boundaries as these are the regions in which all classical methods fail.

We feel that only the use of robust methods can overcome the matching problems that every method encounters at the occluding contours. In [9], a *partial correlation* method is described, based on *robust regression*. It provides good matches at occlusion contours.

In this paper, another partial correlation method is developed, based on *robust location*. It uses a minimum volume estimator (MVE), and is also called the MVE correlation. It can be treated as a robustification of the ZNCC correlation (Zero Mean Cross Correlation), in that sense it is the ZNCC on the *good* parts of the signals.

This paper is organized as follows: after an introduction to related work, we give a short analysis of classical ZNCC correlation from the location view point. Then some fundamental location notions are discussed, followed by a robust location based correlation, which is also called MVE correlation. At the end, some experimental results are shown, including comparisons with two classical methods ZNCC and ZSSD (Zero Mean Sum of Squared differences), Zabih's rank and census [16], and Lan's method RZSSDC (Robust ZSSD with Center point constraint) [9].

2 Related work

In this report, we will describe a *robust* matching algorithm. There are also others papers on matching problem in which some *robust* methods are used. For example, in the paper of Odobez and Bouthemy [13], a *M-estimator* is used to computer the optical-flow, and in [16] and [3], a *non-parametric* transform is used before the correlation. In [9], a *partial correlation* method is described.

2.1 Correlation and image transforms

The visual correspondence of two vectors is usually defined by correlation. There are many different versions of this (see the Table 1 in [1] for instance). Experimental comparisons are reported in [1] and [4].

Among existing methods, we have chosen zero mean normalized cross correlation (ZNCC) because of its good experimental performance with respect of varying lighting conditions:

$$ZNCC(X, X+dX) = \frac{cov(I_1(X), I_2(X+dX))}{\sigma_{I_1(X)}\sigma_{I_2(X+dX)}},$$

with $X = (x, y)$, $dX = (dx, dy)$ the disparity, $cov(I_1, I_2) = \sum_{\Delta \in W} \frac{(I_1(X+\Delta) - \overline{I_1(X)})(I_2(X+dX+\Delta) - \overline{I_2(X+dX)})}{n-1}$ (the covariance between $I_1(X)$ and $I_2(X+dX)$), $\sigma_{I_1} = \sqrt{\sum_{\Delta \in W} \frac{(I_1(X+\Delta) - \overline{I_1(X)})^2}{n-1}}$ (the standard deviation of $I_1(X)$), $\sigma_{I_2} = \sqrt{\sum_{\Delta \in W} \frac{(I_2(X+dX+\Delta) - \overline{I_2(X+dX)})^2}{n-1}}$ (the standard deviation of $I_2(X+dX)$), $\Delta = (u, v) \in W = \{(u, v) | -N \leq u \leq N \text{ and } -M \leq v \leq M\}$, $\overline{I_1(X)}$ and $\overline{I_2(X+dX)}$ the mean gray level values of the two image windows $I_1(X)$ and $I_2(X+dX)$ and $n = (2N+1)(2M+1)$ is the window dimension.

The ZSSD correlation function has also been used:

$$ZSSD(X, X+dX) = \sigma_{I_1(X) - I_2(X+dX)}$$

with $(I_1 - I_2)(\Delta) = I_1(X+\Delta) - I_2(X+dX+\Delta)$, and $\sigma_{I_1(X)-I_2(X+dX)}$ the standard deviation of $I_1(X) - I_2(X+dX)$.

Notice that with ZSSD, we minimize the correlation function while with ZNCC, we maximize it.

Alternative approaches have been proposed based on transformed versions of the original signal. A common standard transformation is the Laplacian of the image. This has a similar effect to normalizing the image as in the ZSSD method.

All of these methods are very sensitive to outliers, e.g. strong noise or the presence of partial occlusions. In order to reduce such effects, some authors limit the influence of the numerical values by only taking into account the sign of the Laplacian [17, 12], or the orientation of the gradient [15]. In these cases, the perturbations of the correlation are in direct ratio to the size of the perturbed area. A similar idea is exploited in the rank and census transform [16] presented in section 2.2. Section 2.3 presents another idea to perform *partial correlation* after having rejected the *occlusion* part [9].

2.2 Rank and census transform

Zabih [16] has developed an approach that relies on local transforms based on *non-parametric* measures, designed to tolerate partial occlusion, which he called *factionalism.* Non-parametric statistics [10] is distinguished by the use of ordering information among data, rather than the data values themselves.

Two local non-parametric transforms were introduced. The *rank transform* is a non-parametric measure of local intensity, while the second one, the *census transform*, is a non-parametric summary of local spatial structure.

Let P be a pixel, $I(P)$ its intensity, and $N(P)$ the set of pixels in its neighborhood. All non-parametric transforms depend on the comparison of $I(P)$ and the intensity of the pixels in $N(P)$. Define $\xi(P, P')$ to be 1 if $I(P') < I(P)$ and 0 otherwise. Non-parametric local transforms depend solely on the set of pixel comparisons, which is the set of ordered pairs :

$$\Xi = \bigcup_{P' \in N(P)} (P', \xi(P, P')).$$

The *rank* transform is defined as the number of pixels in the local region whose intensity is less than the intensity of the center pixel. Formally, the rank transform $R(P)$ is

$$R(P) = card(\{P' \in N(P) | I(P') < I(P)\}).$$

For the rank transform, L_1 correlation (minimization of the sum of absolute values of differences) of the transformed image is used in order to preserve a response that depends linearly on the number of outliers.

The *census* transform $R_\tau(P)$ is a mapping from the local neighborhood of a pixel P to a bit string representing the set of neighboring pixels whose intensity is interior to that of P. Let $N(P) = P \oplus D$ where $\oplus$ is the Minkowski sum and D is a set of displacements, and let $\otimes$ denote concatenation. The census transform can then be specified as:

$$R_\tau(P) = \bigotimes_{[i,j] \in D} \xi(P, P + [i,j]).$$

Pixels of census transformed images are compared for similarity using the Hamming distance (the number of bits that differ) between the two bit strings. The algorithm computes the correspondence by minimizing the Hamming distance after applying the census transform.

2.3 Robust regression based partial correlation

Consider the two windows of Figure 1. They match except for the up-right corners. We want to recover this corrupted part. We assume that locally the signal obeys an affine transformation from one image to the other (see Figure 2):

$$I_2(s,t) = kI_1(s + dx, t + dy) + l + \epsilon$$

where (dx, dy) is the disparity of the point (x, y), $s = x + u, t = y + v$ and $s \in \{x - N \ldots x + N\}, t \in \{y - M \ldots y + M\}$, $2N+1$ and $2M+1$ are the width and the height of the template window respectively and ϵ is a Gaussian image noise.

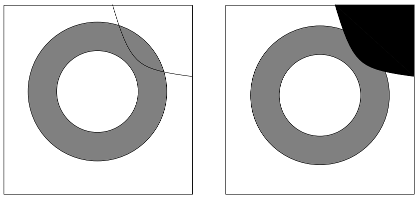


Figure 1: Occlusion.Left: I_1, right: I_2

Such an affine relation is true in the window except for the *occluded* parts.

A simpler model would allow just a shift in intensity, i.e. the scale factor k is set to 1:

$$I_2(s,t) = I_1(s + dx, t + dy) + l + \epsilon.$$

If the center of the window with coordinates (x, y) is *not* occluded, an additional constraint can be stated, which will be called the "*center point constraint*" below. It just states that the center of the window has to satisfy the previous affine or translational constraint:

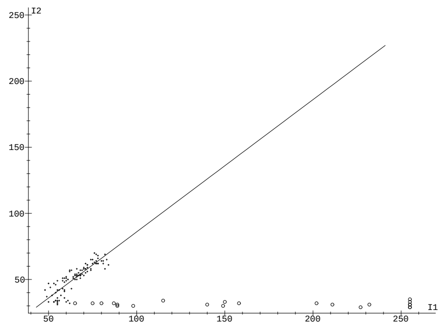


Figure 2: Affine relation between I_1 and I_2, the small circles represent outliers, i.e: pairs of grey levels of pixels mismatched.

$$\begin{aligned} I_2(x,y) &= kI_1(x+dx, y+dy) + l \text{ (affine case)}, \\ I_2(x,y) &= I_1(x+dx, y+dy) + l \text{ (translation case)}. \end{aligned}$$

The occluded parts are found using *robust statistics*. Having found these *occluded* parts, correlation is restricted to the remaining parts of the two windows. We call this technique *partial correlation*. In fact *partial correlation* is a special case of *weighted correlation*, as we leave $\omega(u,v) = 1$ (inlier) or $\omega(u,v) = 0$ (outliers).

For example, defining $D_X(\Delta) = I_1(X+\Delta) - I_2(X+dX+\Delta)$, and $\overline{D_X} = \frac{\sum_{\Delta \in W} D(\Delta)\omega(\Delta)}{\sum_{\Delta \in W} \omega(\Delta)}$, the weighted ZSSD can be expressed as follows:

$$ZSSD_w(X, dX) = \sqrt{\frac{\sum_{\Delta \in W} (D_X(\Delta) - \overline{D_X})^2 \omega(\Delta)}{\sum_{\Delta \in W} \omega(\Delta) - 1}}.$$

3 Analysis of ZNCC: location view point

The function ZNCC of two $n-$dimensional lists $X = (x_i)_{i=1}^n$ and $Y = (y_i)_{i=1}^n$, is defined as:

$$\frac{\sum_{i=1}^n \frac{(x_i - \bar{X})(y_i - \bar{Y})}{n-1}}{\sqrt{\sum_{i=1}^n \frac{(x_i - \bar{X})^2}{n-1} \sum_{i=1}^n \frac{(y_i - \bar{Y})^2}{n-1}}}$$

which can also be expressed as the radio of the covariance $cov(X,Y)$ between X and Y and the geometrical mean of their variances σ_x^2 and σ_y^2.

If we create a list of two-dimensional points $P = (P_i)_{i=1}^n = ((x_i, y_i)^t)_{i=1}^n$, then its center is given by $\bar{P} = (\bar{X}, \bar{Y})^t = (\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n})^t$, its dispersion in terms of the covariance matrix is given by:

$$C = \frac{\sum_{i=1}^n (P_i - \bar{P})(P_i - \bar{P})^t}{n-1} = \begin{pmatrix} \sigma_x^2 & cov(x,y) \\ cov(x,y) & \sigma_y^2 \end{pmatrix}$$

with $\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n-1}}$, $\sigma_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{Y})^2}{n-1}}$, $cov(x,y) = \frac{\sum_{i=1}^{n}(x_i-\bar{X})(y_i-\bar{Y})}{n-1}$, and $ZNCC(X,Y) = \frac{cov(x,y)}{\sigma_x\sigma_y} = \frac{C_{12}}{\sqrt{C_{11}C_{22}}}$, where C_{ij} is the element of C in the i^{th} row and the j^{th} column.

We see that the ZNCC function of $X = (x_i)$ and $Y = (y_i)$ can be computed from the dispersion of $P = (P_i) = (x_i, y_i)^t$.

This procedure is composed of three steps:

1. Estimation of the data center.
2. Estimation of the covariance matrix.
3. Estimation of the correlation coefficient by computing the ratio of the two variable covariance and the geometrical mean of their variances.

In the partial occlusion case, there are some outliers. The estimation of the center and the covariance matrix should be robust so that the estimation of the correlation coefficient is reliable. In the next section, we discuss the robust estimation of location and dispersion.

4 Robust estimation of location and dispersion

In this section, we introduce the notion of a location estimator, consider properties as translation invariance, permutation invariance and affine invariance, and discuss the notion of breakdown point. At the end, we introduce some robust estimators, and give more details of MVE, which is affine invariant. This estimator is applied to deduce a correlation function, which is robust to partial occlusions (section 5).

4.1 Estimation of location and dispersion

We discuss the estimation of the "center" of a set of points, where all variables are treated in the same way. We are also interested in the data dispersion around the center, which is expressed by the covariance matrix of this set of points.

Suppose that we have a set of data:

$$\begin{aligned}\mathbf{X} &= \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \\ &= \{(\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1p})^t, \ldots, (\mathbf{x}_{n1}, \mathbf{x}_{n2}, \ldots, \mathbf{x}_{np})^t\},\end{aligned}$$

of n p-dimensional points. We have to estimate its center, and also its dispersion around the center, expressed by a covariance matrix C.

More exactly, we are looking for an estimator $\mathbf{T}$, such that $\mathbf{e}_i = \mathbf{x}_i - \mathbf{T}(\mathbf{X})$ follows a hypothesized distribution and has some useful properties. If some outliers among the $\mathbf{e}_i$ do not perturb the result, the estimator is said to be *robust*.

4.2 Some definitions and an example

We introduce some properties of robust estimators, including the notions of invariance and breakdown point.

Definition 1 *A location estimator* $\mathbf{T}$ *is called translation invariant, if and only if for any vector* b*,*

$$\mathbf{T}(\mathbf{x}_1 + \mathbf{b}, \ldots, \mathbf{x}_n + \mathbf{b}) = \mathbf{T}(\mathbf{x}_1, \ldots, \mathbf{x}_n) + \mathbf{b}.$$

Definition 2 *A location estimator* $\mathbf{T}$ *is called permutation invariant if for each permutation* π *on* $\{1, 2, \ldots, n\}$*,*

$$\mathbf{T}(\mathbf{x}_{\pi(1)}, \ldots, \mathbf{x}_{\pi(1)}) = \mathbf{T}(\mathbf{x}_1, \ldots, \mathbf{x}_n).$$

The sample mean $\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$, has these properties and also minimizes $\sum_{i=1}^{n} ||\mathbf{x}_i - \mathbf{T}||^2$, where $||\ldots||$ is the Euclidean norm. However, it is not robust, as a single bad point can totally corrupt the result.

Now consider the notion of *robustness* for a location estimator. Suppose the sample is corrupted by replacing some m data points with arbitrary values, and define the maximal bias:

$$bias(m; \mathbf{T}, \mathbf{X}) = \sup_{\mathbf{X}'} ||\mathbf{T}(\mathbf{X}') - \mathbf{T}(\mathbf{X})||$$

and the breakdown point

$$\epsilon_n^\star(\mathbf{T}, \mathbf{X}) = \min \{m/n; bias(m; \mathbf{T}, \mathbf{X}) \text{ is infinity}\}.$$

This is the minimal outlier percentage that can arbitrarily damage $\mathbf{T}$.

The sample mean has its breakdown point at $\frac{1}{n}$. We often consider the breakdown point for $n \to \infty$, so it has an asymptotic breakdown point of 0%.

It is clear that a translation invariant estimator can not have a breakdown point greater than 50%: if the outliers happened to correspond to a translation of good data, a good choice of $\mathbf{T}$ would be impossible.

The estimators with a breakdown point at 50% fall into two families: those which are affine invariant and those which are not. In many situations, we hope to have an estimator which commutes with affine transformations. $\mathbf{T}$ is said to be affine invariant if and only if:

$$\mathbf{T}(\mathbf{x}_1\mathbf{A} + \mathbf{b}, \ldots, \mathbf{x}_n\mathbf{A} + \mathbf{b}) = \mathbf{T}(\mathbf{x}_1, \ldots, \mathbf{x}_n)\mathbf{A} + \mathbf{b}$$

for any vector $\mathbf{b}$ and any regular matrix $\mathbf{A}$. For example, the sample mean is such an estimator. Note that all translations and permutations are affine, so affine invariance implies translation and permutation invariance.

4.3 MVE

Rousseeuw [14] has introduced an affine invariant robust estimator, which has the theoretically maximal breakdown point, defined as: the "center of the ellipse of minimal volume containing at least h points of X" denoted as MVE.

Here h can be taken as $[\frac{n}{2}] + 1$. This is called Minimal Volume Ellipse or MVE. The ellipse itself gives an estimate of the covariance matrix, multiplied by a suitable scale factor.

The affine invariance of this estimator follows from the fact that the image of an ellipse by an affine transformation $\mathbf{x} \rightarrow \mathbf{A}\mathbf{x} + \mathbf{b}$ is still an ellipse, with the volume scaled by $|det\mathbf{A}|$. Since $|det\mathbf{A}|$ is constant when the data change, the relative dimension of the ellipse does not change under affine transformations.

It is not practical to consider all possible subdivisions of data and compute the volume of the smallest ellipse around them. As the LMedS estimator, we use a random sampling technique to implement the estimator.

We choose a set of $p+1$ different observations, indexed by $\mathbf{J} = \{i_1, \ldots, i_{p+1}\}$. For this subset, we determine the mean and the corresponding covariance matrix by:

$$\overline{\mathbf{X_J}} = \frac{1}{p+1} \sum_{i \in \mathbf{J}} \mathbf{x}_i$$

and

$$\mathbf{C_J} = \frac{1}{p} \sum_{i \in \mathbf{J}} (\mathbf{x}_i - \overline{\mathbf{x_J}})(\mathbf{x}_i - \overline{\mathbf{x_J}})^t$$

where $\mathbf{C_J}$ is regular if $\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_{p+1}}$ are in general position. By the definition of median, the ellipse containing half the data corresponds to the matrix $m_{\mathbf{J}}^2 \mathbf{C_J}$, with $m_{\mathbf{J}}^2 = \underset{i}{\operatorname{med}} \ (\mathbf{x}_i - \overline{\mathbf{X_J}})^t \mathbf{C_J}^{-1} (\mathbf{x}_i - \overline{\mathbf{X_J}})$. It suffices to see that half of the $\mathbf{x}_i$ verify: $(\mathbf{x}_i - \overline{\mathbf{X_J}})^t (m_{\mathbf{J}}^2 \mathbf{C_J})^{-1} (\mathbf{x}_i - \overline{\mathbf{X_J}}) = \frac{(\mathbf{x}_i - \overline{\mathbf{X_J}})^t \mathbf{C_J}^{-1} (\mathbf{x}_i - \overline{\mathbf{X_J}})}{m_{\mathbf{J}}^2} \leq 1$.

The volume of the ellipse is proportional to:

$$\sqrt{det(m_{\mathbf{J}}^2 C_{\mathbf{J}})} = \sqrt{det(\mathbf{C_J})} m_{\mathbf{J}}^p. \tag{1}$$

We repeat this process for several $\mathbf{J}$ and we take the $\mathbf{J}$ which gives the minimal volume in equation (1).

Then we compute:

$$\mathbf{T}(\mathbf{X}) = \overline{\mathbf{X_J}} \qquad \text{and} \qquad \mathbf{C}(\mathbf{X}) = (\chi^2_{p,0.5})^{-1} m_{\mathbf{J}}^2 \mathbf{C_J}$$

where $\chi^2_{p,0.5}$ is the median of a chi-square distribution with p degrees of freedom.

The estimated MVE can be used as an initial solution in refinement. Similarly to the case of weighted least squares estimation, we give a weight to each observation by the rule:

$$\omega_i = \begin{cases} 1 & \text{if } (\mathbf{x}_i - \mathbf{T}(\mathbf{X}))^t \mathbf{C}(\mathbf{X})^{-1} (\mathbf{x}_i - \mathbf{T}(\mathbf{X})) \leq c \\ 0 & \text{else} \end{cases}$$

where C can take the value $\chi^2_{p,0.975}$ which means that we reject a point when we have 97.5% confidence that it is an outlier.

We use the weighted mean estimator :

$$\mathbf{T}_1(\mathbf{X}) = \frac{\sum_{i=1}^n \omega_i \mathbf{x}_i}{\sum_{i=1}^n \omega_i}$$

and

$$\mathbf{C}_1(\mathbf{X}) = \frac{\sum_{i=1}^{n} \omega_i (\mathbf{x}_i - \mathbf{T}_1(\mathbf{X}))(\mathbf{x}_i - \mathbf{T}_1(\mathbf{X}))^t}{\sum_{i=1}^{n} \omega_i - 1}$$

which means simply that the classical computation is performed on the observations whose weights are 1.

4.4 Application of MVE estimator

The MVE estimator has been already used in vision, as a classification method. For example, in [8], it is used for thresholding, for analysis of Hough space, and for image segmentation. We will use it for the image matching problem in this paper (section 5).

5 Robust location based partial correlation

We begin with a discussion of correlation and location, deduce the procedure for partial correlation, and finally, we introduce the *main points constraint* which can be used in the matching of two images.

5.1 Analysis of correlation and location

Suppose we have a list of n p-dimensional points: $\mathbf{X} = (\mathbf{x}_i)_{i=1}^n$, where $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots \mathbf{x}_{ip})^t$. We look for an estimator $\mathbf{T} = \mathbf{T}(\mathbf{X})$, such that:

$$\mathbf{x}_i = \mathbf{T} + \mathbf{e}_i \tag{2}$$

where $\mathbf{e}_i = (\mathbf{e}_{i1} \dots \mathbf{e}_{ip})$ are some vectorial random variables.

The location problem is the search for such an estimator. Under the assumption that the $\mathbf{e}_i$ have an identical distribution, the correlation coefficient between $\mathbf{e}_{ij}$ and $\mathbf{e}_{ik}$ depends only on j and k. It can be computed by the aid of $\mathbf{e}_i$'s covariance matrix, as the ratio of the covariance between $\mathbf{e}_{ij}$ and $\mathbf{e}_{ik}$ and the geometrical mean of $\mathbf{e}_{ij}$ and $\mathbf{e}_{ik}$'s variances.

In image matching problem, $p = 2$, $\mathbf{x}_i = (I_{1i}, I_{2i})$ represents a pair of corresponding intensities, $\mathbf{T}$ represents the "center" of $(\mathbf{x}_i)_{i=1}^n$ and it can be estimated as the gravity of points $\mathbf{x}_i$ representing matched pixels, $\mathbf{e}_i = \mathbf{x}_i - T$ represents the "joined" texture of I_1 and I_2. The correlation coefficient between $\mathbf{e}_{i1}$ and $\mathbf{e}_{i2}$ reflexes the association between I_1 and I_2, and it can be estimated by estimating the covariance matrix of $\mathbf{e}_i$.

5.2 Dealing with the partial occlusion problem

Most correlation techniques have difficulties near disparity discontinuities, or in places where highlights occur, as the window under consideration is locally partially but severely corrupted which occurs at the boundaries of objects (Figure 1).

The pixels in such a region represent scene elements from two distinct intensity populations. Some of the pixels correspond to the template under consideration and some to other parts of the scene.

As we already mentioned, this leads to a problem for many correlation techniques, since the standard correlation techniques are usually based on standard statistics methods, which are best suited to a single population. We shall call this phenomenon *partial occlusion* even though it might have other physical causes as highlights, and we propose the *partial correlation* idea to overcome it.

5.3 Procedure of partial correlation

When partial occlusions occur, relation (2) holds only for the non-occluded part. So we should compute the correlation measure only on this part.

We propose the partial correlation procedure as follows:

1. Estimate robustly T and the associated covariance matrix C.
2. Find the occluded part according to a statistical test.
3. Compute the correlation measure on the non occluded part.

We explain these steps in detail by some examples:

1. We can take T as the median of each coordinate estimator or a more sophisticated estimator such as MVE.
2. Estimate the distribution of the errors, and decide if the errors follow this distribution according to a statistical test. If an error does not follow this distribution, the associated datum is considered to be an outlier.

   Example: In the MVE estimator, $(\mathbf{x}_i - \mathbf{T})^t \mathbf{C}^{-1}(\mathbf{x}_i - \mathbf{T})$ is a χ^2 random variable with p degrees of freedom, if all the $\mathbf{e}_i$ are centered Gaussian. If the value of this expression exceeds a certain threshold ($\chi^2_{p,0.975}$ for example), the datum $\mathbf{x}_i$ is declared to be an outlier.
3. Partial correlation can be achieved by binary weighted correlation. For example, if we leave $\omega_i = 1$ for good data and $\omega_i = 0$ for bad data, partial ZNCC can be expressed in terms of weighted correlation as follows:

$$ZNCC_p(X,Y) = \frac{\frac{\sum_{i=1}^n (x_i-\bar{x})(y_i-\bar{y})\omega_i}{\sum_{i=1}^n \omega_i - 1}}{\sqrt{\frac{\sum_{i=1}^n (x_i-\bar{x})^2\omega_i}{\sum_{i=1}^n \omega_i - 1}}\sqrt{\frac{\sum_{i=1}^n (y_i-\bar{y})^2\omega_i}{\sum_{i=1}^n \omega_i - 1}}}$$

   where $\bar{x}$ (respectively $\bar{y}$) is the ω-weighted means of x_i (respectively $\bar{y}$).

5.4 Main point constraint

As in the implementation of the LMedS estimator, let $p+1=3$ be the parameter number and ϵ be the outlier percentage. Then, to have a confidence Q that at least one random sample contains only inliers, $\frac{\log Q}{\log(1-(1-\epsilon)^{p+1})}$ samples suffice [14]. If we put $\epsilon = 0.5$, (the maximum value allowed) and $Q = 0.01$, then we have $\frac{\log Q}{\log(1-(1-\epsilon)^{p+1})} = 34.49$. So the sample number is 35.

In our experiments, we divide the 3×3 window around the pixel to be matched into four parts, as follows:

3	4	4
3	c	1
2	2	1

Numbers in this board indicate different regions and c denotes the pixel to be matched. We suppose that the pixel to be matched is not occluded, and the two pixels in at least one region are not occluded neither.

This is called the *main point constraint*. The points around are called *main points*.

Using this constraint, it suffices to take c with four groups of pixels in the sampling procedure, and the number of samples is reduced from 35 to 4.

The advantages of this constraint are 1)it reduces the sample number and hence the computation time; 2)it incorporates the fact that adjacent neighbors of the pixel to be matched are either both likely to be occluded or both escape occlusion. Its shortcoming is that the location estimation is more dependent on the main points.

5.5 Estimating correlation between two image windows

In case of image matching, we take $\mathbf{x}_i = (I_{1i}, I_{2i})^t$, where I_{1i} (I_{2i}) are intensities of the i^{th} point of the first (second) image window respectively.

Let $\mathbf{C}_1(\mathbf{X})$ be the covariance matrix of $\mathbf{x}_i$ estimated by the MVE. Defining

$$\mathbf{C}_1(X) = \begin{pmatrix} \sigma_1^2 & cov(I_1, I_2) \\ cov(I_1, I_2) & \sigma_2^2 \end{pmatrix},$$

we estimate the correlation coefficient between I_1 and I_2 as $\frac{cov(I_1,I_2)}{\sigma_1\sigma_2}$. This estimate is the generalization of the classical correlation coefficient if $\mathbf{C}_1(\mathbf{X})$ is replaced by the classical covariance matrix, which is computed as:

$$\begin{pmatrix} \frac{\sum_{i=1}^n (I_{1i}-\overline{I_1})^2}{n-1} & \frac{\sum_{i=1}^n (I_{1i}-\overline{I_1})(I_{2i}-\overline{I_2})}{n-1} \\ \frac{\sum_{i=1}^n (I_{1i}-\overline{I_1})(I_{2i}-\overline{I_2})}{n-1} & \frac{\sum_{i=1}^n (I_{2i}-\overline{I_2})^2}{n-1} \end{pmatrix}$$

with $\overline{I_1} = \frac{\sum_{i=1}^n I_{1i}}{n}$ and $\overline{I_2} = \frac{\sum_{i=1}^n I_{2i}}{n}$ are the grey level means of the two windows. The classical correlation coefficient, often called ZNCC, is a frequently used correlation function in vision for the matching problem.

A real example is shown in Figure 3, where each point represents a pair of corresponding pixel grey levels (I_{1i}, I_{2i}). The ellipse is similar to the uncertainty ellipse of the data set.

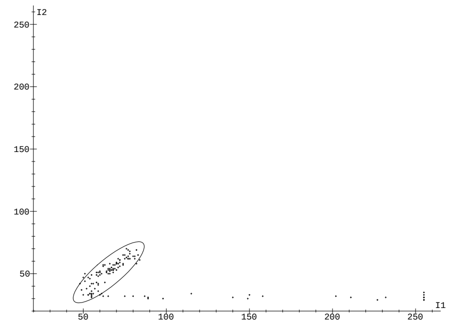


Figure 3: Robust location by MVE estimator of (I_1, I_2). The points outside the ellipse represent pairs of points that are not matched.

6 Experimental results

6.1 Constraints used

Select a template of fixed size in the first image. Suppose that there are several candidate matches of the same size. Below we use the *region of interest* constraint and the *epipolar constraint* to reduce the number of the candidates [1].

Roughly say, the *region of interest* constraint is that the match of each pixel in an image should be contained in the *region of interest* (ROI), which is defined as $\{(x', y')|x + dx_{min} \leq x' \leq x + dx_{max}, y + dy_{min} \leq y' \leq y + dy_{max}\}$, (x, y) is the pixel chosen in the first image and $dx_{min}, dx_{max}, dy_{min}, dy_{max}$ are the limits of disparities in the two directions x and y. The *epipolar constraint* is that the match of each pixel in an image should be on its epipolar line in another image.

These two constraints are displayed in the Figure 4: the window candidate has to be centered on the corresponding epipolar line and stay within the region of interest.

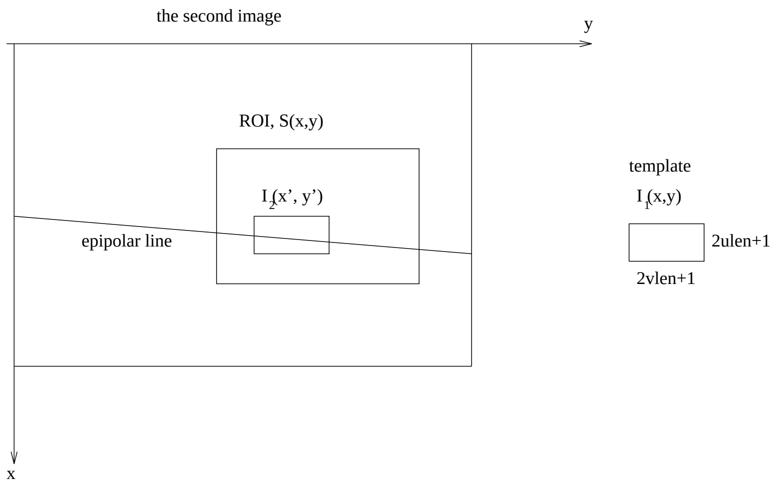
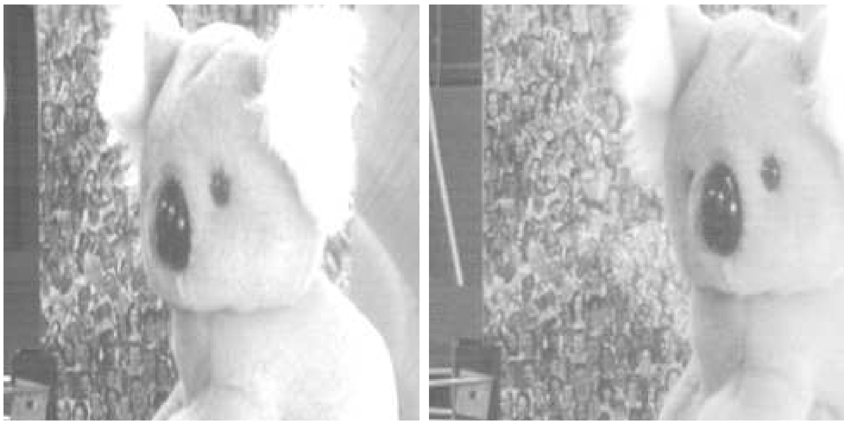


Figure 4: Epipolar constraint and region of interest constraint.

6.2 Statistical results

For comparison with ground truth, experiments were conducted on planar patches. In such cases the exact match can be computed by estimating the homography mapping from the matched targets between two images. We have tested the MVE correlation, and compared it with the ZNCC method, the partial regression based method using the translation relation and the center point constraint (RZSSDC) [9], which stands for *Robust ZSSD with Center point constraint*, the standard ZSSD, the rank and the census methods [16]. All the methods ZNCC, ZSSD, rank, census, RZSSDC have been discussed in section 2.

The results for two pair of images called *Benetton* (Figure 5) and *World* (Figure 6) are displayed in this report.



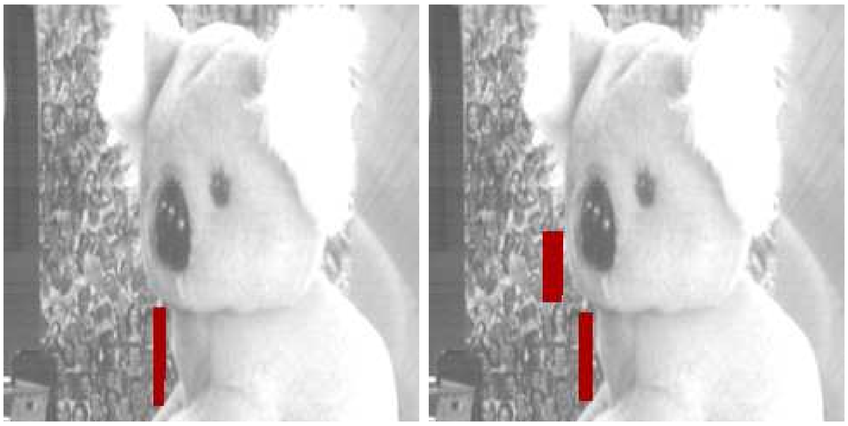
(a) (b)

Figure 5: Stereo pair 1: *Benetton*

(a) (b)

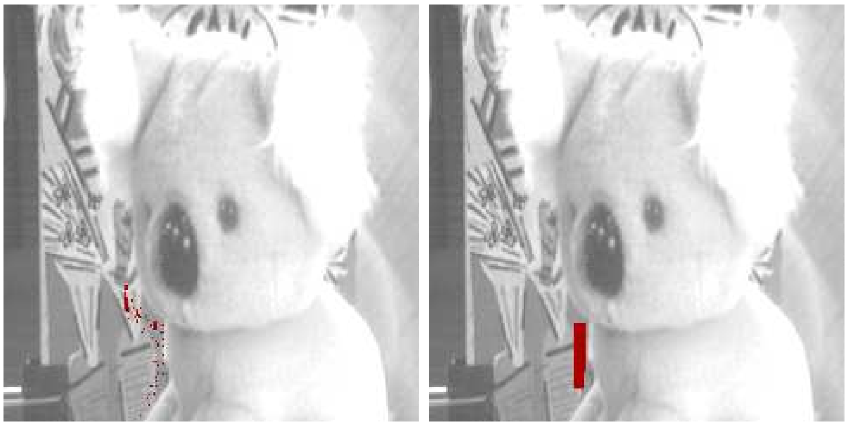
Figure 6: Stereo pair 2: *World*

Both images have a planar background but the background textures are different. Two tests on points near occlusion contours were designed for each pair. Figure 7 and 8 displays respectively two sets of points for the *Benetton* (Figure 5 (a)) and *World* (Figure 6 (a)) background for which the corresponding points were sought.



(a) (b)

Figure 7: Background *Benetton*. Points selected (in black) in Figure 5 (a). (a): example corresponding to Table 1; (b): example corresponding to Table 2.



(a) (b)

Figure 8: Background *World*. Points selected in Figure 6 (a). (a): example corresponding to Table 3; (b): example corresponding to Table 4.

The different methods were run on these four tests. The candidates were constrained to stay on the epipolar line and to have a disparity of less than 200. For each method, we compare the results found with the exact value provided by the homography.

The results are reported in the Tables 1, 2. 3 and 4. For each method (i.e: rank, census, RZSSDC, ZSSD, MVE and ZNCC) are indicated:
- the number of accurate matches: up to one pixel error (good match),
- the number of matches in a distance between one and two pixels error (near miss),
- the number of matches at a distance of two to three pixels (bad matches),
- the matches which lie more than three pixels away from their exact position (false matches).

	0–1	1–2	2–3	$3-\infty$
rank	246	96	0	57
census	305	63	0	31
RZSSDC	316	63	0	20
ZSSD	188	37	0	174
MVE	270	79	0	50
ZNCC	183	37	0	179

Table 1: First example on image *Benetton* (see text).

	0–1	1–2	2–3	$3-\infty$
rank	713	98	0	89
census	773	83	0	44
RZSSDC	796	74	0	30
ZSSD	658	38	3	201
MVE	730	108	0	56
ZNCC	656	38	1	205

Table 2: Second example on image *Benetton* (see text).

	0–1	1–2	2–3	$3-\infty$
rank	77	14	0	49
census	89	12	0	39
RZSSDC	118	10	0	12
ZSSD	56	0	0	84
MVE	94	10	1	35
ZNCC	56	0	0	84

Table 3: First example on image *World* (see text).

	0–1	1–2	2–3	$3-\infty$
rank	129	15	0	122
census	185	6	0	75
RZSSDC	234	29	2	1
ZSSD	71	1	0	194
MVE	173	49	7	37
ZNCC	78	0	0	188

Table 4: Second example on image *World* (see text).

From these examples, we see that the MVE correlation performs better than the ZNCC correlation for points near occlusion contours.

The results on general points are displayed below, which show that a compromise between *similarity* and *completeness* is needed [9]. For the image pair *Benetton*, MVE correlation gives less false matches than ZNCC correlation. Comparing to RZSSDC, we see that MVE gives less good results for points near occlusion contours, but better results for general points.

7 Conclusion and future work

Occlusions carry important information in the visual correspondence problem, because they indicate disparity discontinuities and produce useful cues for the shape and location of objects. However, occlusions show the limitations of most of standard correlation methods.

In this paper, we discussed the occlusion problem and proposed a new robust approach called *partial correlation* to overcome it. We compared our method with the standard correlation methods ZNCC

	0–1	1–2	2–3	3–∞
rank	12500	2053	3	214
census	12885	1745	0	140
RZSSDC	12585	1930	4	251
ZSSD	12615	1580	0	575
MVE	12378	2183	40	169
ZNCC	12771	1550	0	449

Table 5: An example for the general case on image *Benetton* (see text).

	0–1	1–2	2–3	3–∞
rank	11946	2512	299	968
census	12960	2237	105	423
RZSSDC	11845	1985	193	1702
ZSSD	13140	1941	116	528
MVE	11102	2512	414	1697
ZNCC	13216	1970	80	459

Table 6: An example for the general case on image *World* (see text).

and ZSSD, Zabih's correlation method rank and census and Lan's Method RZSSDC. Experiments show that our method works better than classical correlation methods when occlusions arise. However, when no occlusions are present, our method performs less well.

We think that the *partial correlation* represents only the beginning of the solution to the problem. The fundamental limitation of this technique is the necessity for a tradeoff between the *similarity* and the *completeness* criteria : *Given a template, and several candidates, which one is the most similar to the template ?*

The most similar candidate may have few active pixels hence not be very complete. The straightforward solution could be to take the one which is more similar when occlusions occur and the one which is more complete otherwise. This is an important issue not only in the problem of stereo matching by correlation, but also in the more general template matching problems like the recognition. A possible perspective is the application of *partial correlation* technique to the recognition problem, where partial occlusion often occurs.

References

[1] P. Aschwanden and W. Guggenbühl. Experimental results from a comparative study on correlation-type registration algorithms. In Förstner and Ruwiedel, editors, *Robust Computer Vision*, pages 268–282. Wichmann, 1992.

[2] D.H. Ballard and C.M. Brown. *Computer Vision.* Prentice Hall, 1982.

[3] D. N. Bhat and S. K. Nayar. Ordinal measure for visual correspondence. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*, pages 351 – 357, San Francisco, California, June 1996.

[4] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, 1990.

[5] D. Geiger, B. Ladendorf, and A. Yuille. Occlusions and binocular stereo. In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 425–433. Springer Verlag, 1992.

[6] B.K.P. Horn. *Robot Vision.* The MIT Press, 1986.

[7] S.S. Intille and A.F. Bobick. Disparity-space images and large occlusion stereo. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 179–186. Springer-Verlag, 1994.

[8] J.M. Jolion, P. Meer, and S. Bataouche. Robust clustering with applications in computer vision. *PAMI*, 13(8):791–802, August 1991.

[9] Z. D. Lan, R. Mohr, and P. Remagnino. Robust matching by partial correlation. In *Proceedings of the sixth British Machine Vision Conference, Birmingham, England*, pages 651–660, September 1995.

[10] E. L. Lehman. *Nonparametrics: statistical methods based on ranks.* Holden-Day, 1975.

[11] D. Marr. *Vision.* W.H. Freeman and Company, San Francisco, California, USA, 1982.

[12] H.K. Nishihara. PRISM, a pratical real-time imaging stereo matcher. Technical Report Technical Report A.I. Memo 780, Massachusetts Institute of Technology, 1984.

[13] J.M. Odobez and P. Bouthemy. Estimation robuste multi-échelle de modèles paramétrés de mouvement sur des scènes complexes. *International Journal of Computer Vision*, 11:419–430, 1994.

[14] P.J. Rousseeuw and A.M. Leroy. *Robust regression and outlier detection*, volume XIV of *Wiley.* J.Wiley and Sons, New York, 1987.

[15] P. Seitz. Using local orientational information as image primitive for robust object recognition. In *SPIE proceedings*, pages 1630–1639, 1989.

[16] R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondance. In *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 151–158. Springer-Verlag, May 1994.

[17] J. Zhao. *Extraction d'information tri-dimensionnelle par stéréovision.* Thèse de doctorat, Université Paul Sabatier, Toulouse, July 1989.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399