INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Non-parametric Invariants and Application to Matching

Zhong-Dan LAN and Roger MOHR

N° 3246

September 5, 1997

THÈME 3

Rapport de recherche

ISSN 0249-6399

Non-parametric Invariants and Application to Matching

Zhong-Dan LAN and Roger MOHR

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

Rapport de recherche n° 3246 — September 5, 1997 — 23 pages

Abstract: This report present a non parametric matching method based on non parametric invariants to geometry and intensity transforms of images. The use of invariant makes it possible to match images under large geometric transforms. The use of non parametric measures makes it robust to partial occlusion. An algorithm to obtain sub-pixel precision is also shown. Results on real images validate the approach.

Key-words: matching, invariants, interest points, partial occlusion, non-parametric, robustness, precision

(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

Invariants non paramétriques et applications aux appariements

Résumé : Ce rapport présent une méthode d'appariement non paramétriques qui se base sur des mesures non paramétriques invariant aux transformations géométriques et photogrammétriques. Cette méthode est robuste au bruit, aux effets non linéaires d'intensités et aux occultations partielles. Des résultats expérimentaux sur des images réelles valident cette approche et montrent l'amélioration significative par rapport à la méthode différentielle [25].

Mots-clé : appariement, invariants, points d'intérêt, occultation partielle, non paramétrique, robustess, précision

Contents

1 Introduction

Matching is a major issue in computer vision [7, 4, 12, 20], as it is necessary for stereo-vision, motion estimation, object recognition etc.

1.1 Existing matching methods

It is very difficult to give an overview for such a widely spread subject as matching. We limit ourselves to the 2D images matching case.

1.1.1 Contour based methods

Some authors propose to use contours as they contain significant information. One of the earliest works can be found in [3], where straight line segments are used as matching primitive in three views case and some heuristic constraints are used as well as the similarity measure between the segments. More sophisticate methods are proposed by Horaud [11], where relations are built on contours, and a search for maximal cliques is applied. However, its time complexity is very high, and therefore additional constraints like the epipolar constraint have to be added in order to reduce the run time.

As indicated in [3], all contour based methods can only be applied if the segmentation problem is well solved, this hypothesis is however only realistic for images of simple scene.

1.1.2 Area based methods

The correlation technique [2] has been used for a long time [32, 5, 16, 18]. In order to obtain satisfactory results, the corresponding windows should be approximately superposed by a simple translation. In case of non-trivial image rotation, the number of false matches increases. A method to correct the false matches and to get precise matches can be found in [17].

Other methods such as Hausdorff distance [14], frequential method [6, 28] can also be used, in case of small image rotation.

1.1.3 Dealing with rotation

The above techniques as correlation, Hausdorff distance, frequential method are suitable for of translational case only. A tool for matching images under more general motion as rotation is required. This is necessary in case of auto-calibration which implies different rotations with non-parallel axes which leads to a great geometrical change between the images of a same scene. Thus the object identification has to be stable to rigid displacement (rotation and translation) between the images.

The matching under these general conditions is difficult and thus it will be performed on interest points [10] only as they are local and the grey levels in their neighborhood contain most information which makes their identification easier.

To obtain insensibility to rotation and scale changes, the translation-base techniques are used in several directions and at several scales. Hu [13] suggests applying the correlation in several directions and Wu [31] proposed using Gabor filters in several directions and at several scales. Rucklidge [24] suggests applying the Hausdorff distance to the task of locating an affine transformation, using a rasterized approach to the search.

In order to avoid calculation of several values, Schmid [25] proposed to use local measures which are invariant to rotation and she proposed to use a multi-scale approach to deal with the scale change.

1.2 Partial occlusion: translational case

In this section, we present some partial and non-parametric correlations to deal with the partial occlusion problem; other methods, as the generalized Hausdorff distance [14] are also possible.

Most correlation techniques have difficulties near disparity discontinuities, or in places where highlights occur, as the window under consideration is locally partially but hardly corrupted in a small area. of objects. Therefore in such a case the pixels in a local region represent scene elements from two distinct intensity populations. Some of the pixels correspond to the template under consideration and some from another 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 call this phenomenon *partial occlusion* even it might have other physical sources as highlights,

1.2.1 Partial correlation

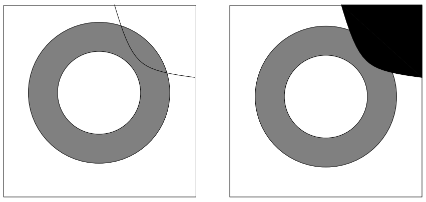


Figure 1: Occlusion.Left: I_1, right: I_2.

Partial correlation idea has been proposed to overcome this difficulty. Consider the two windows of the Figure 1, they match except for the right-up part in the image I_1 and I_2. We have to recover this corrupted part. We assume that locally the signal obeys an affine relation from on 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.

Such an affine relation is true in the window except for the *occluded* parts (as the right-up part indicated in the Figure 1).

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

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

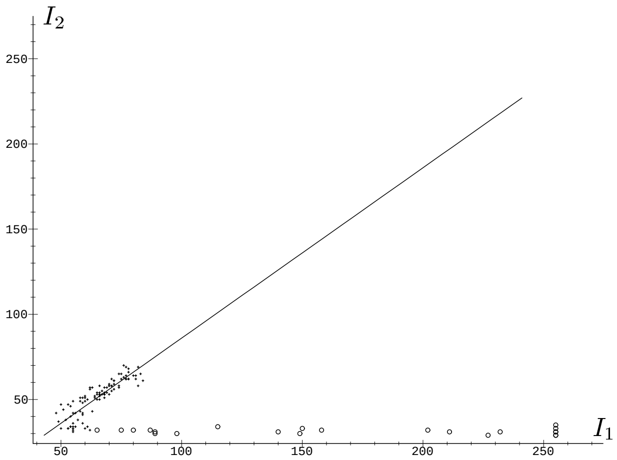


Figure 2: Affine relation between I_1 and I_2, the small circles are outliers.

The occluded part is found using *the robust statistics* [23]. Having found the *occluded* parts, the correlation is restricted on 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 $\omega(u,v) = 1$ (inliers) or $\omega(u,v) = 0$ (outliers).

For example, the weighted $ZSSD$ can be expressed as follows :

$$ZSSD_w(X, dX) = \sqrt{\frac{\sum_{\Delta \in W}(((I_1(X+\Delta) - \bar{I}_1(X)) - (I_2(X+\Delta) - \bar{I}_2(X)))^2 weight(\Delta))}{\sum_{\Delta \in W} weight(\Delta)}}. \quad (1)$$

Another partial correlation, instead of adjusting a line to the intensities pairs, looks for an ellipse of smallest volume what contains at least half of the data, and uses this ellipse to eliminate the outliers among the data. The correlation coefficient is estimated using the remained data. Interested readers can refer to [18].

Partial correlations have some successes near the disparity discontinuity. However, they use pixel-to-pixel correspondence to eliminate the outliers, this correspondence exists in translational cases, and loses in rotation case. Another drawback is that they are based on some linear models, and therefore have some difficulty to handle non linear effects.

1.2.2 Non Parametric correlations

Non parametric correlations [32, 5] do not eliminate the outliers according to a parametric model, instead, they use only non parametric models, which are less sensitive to outliers. In addition, they can handle non linear effects because non hypothesis of linearity are used.

Rank and census transform [32] Zabih has developed an approach which relies on local transforms based on non-parametric measures. They are designed to tolerate *factionalism*. Non parametric statistics [19] is distinguished by the use of ordering information among data, rather than the data values themselves.

Two local non-parametric local transforms were introduced. The first one, called *rank transform*, is a non-parametric measure of local intensity. And the second one, called *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 neighbor. All non parametric transforms depend upon 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. The 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

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

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

The *census* transform $R_\tau(P)$ is a mapping from the local neighbor surrounding a pixel P to a bit string representing the set of neighboring pixels whose intensity is less than 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:

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

Two pixels of census transformed image are computed for similarity using the Hamming distance, i.e. the number of bits that differ in two bit strings. The algorithm computes the correspondence by minimizing the Hamming distance after applying the census transform.

An example These two transforms tolerate the outliers, as displayed in the following. Consider a window of dimension 3×3 of an image whose intensities are :

$$\begin{array}{ccc} 127 & 127 & 129 \\ 126 & 128 & 129 \\ 127 & 131 & A \end{array}$$

where $0 \leq A < 256$.

Consider the set of effects of several parametric and non parametric measures, computed at the window center, when A varies over its 256 possibles values.

The mean of this window varies from 114 to 142, and the variance ranges from 2 to 1823. These non-parametric measures exhibit continuous variation over a substantial range as A changes.

Non parametric transforms are more stable. All the elements of Ξ except one will remain fixed as A changes. Ξ will be:

$$\begin{array}{ccc} 1 & 1 & 0 \\ 1 & & 0 \\ 1 & 0 & a \end{array}$$

where a is 1 if $A < 128$, 0 otherwise. The census transform results $\{1,1,0,1,0,1,0,a\}$ and the rank transform will give 5 if $A < 128$, otherwise 4 .

This comparison shows the tolerance that non parametric measures have for factionalism. A minority of pixels can have a very different value, but the effect on the rank and census transforms is limited by the size of the minority.

Bhat's non parametric correlation Bhat [5] ranks the intensity values in each window and gets a permutation. Using a new distance between the permutations, a new correlation coefficient is obtained.

For a set of window intensity values $(I_1^i)_{i=1}^n$ and $(I_2^i)_{i=1}^n$. Let π_1^i (and π_2^i) be the ranks of I_1^i (et I_2^i) among the I_1 (et I_2) data.

Define a composite permutation as follows: $s^i = \pi_2((\pi_1^{-1})(i))$, where π_1^{-1} denotes the inverse permutation of π_1. Informally, s^i is le rank of the pixels in I_2 which corresponds to the pixel with rank i in I_1.

Under perfect positive correlation, s should be equal to the identical permutation given by $u = (1, 2, \dots n)$.

By defining a distance measure between s and u we in turn obtain a notion of distance between π_1 and π_2.

The deviation d_m^i at each s_i is defined as the number of s_j $(j = 1 \dots i)$ bigger than i. Formally, $d_m^i = \sum_{j=1}^{i} 1_{[i+1,+\infty[} s^j$, where 1 is the indicator function.

Now, a correlation measure $\kappa = \kappa(I_1, I_2)$ is defined as:

$$\kappa(I_1, I_2) = 1 - \frac{2 \max_{i=1}^n d_m^i}{[\frac{n}{2}]}$$

If I_1 and I_2 are perfectly correlated, then $\kappa = 1$. κ has the following properties :

- it is independent of scaling and shift of the intensity values.
- it is a normalized measure : $-1 \leq \kappa \leq 1$.
- it is symmetrical, i.e. $\kappa(I_1, I_2) = \kappa(I_2, I_1)$.
- $\kappa(f(I_1), h(I_2)) = \kappa(I_1, I_2)$ if f and h are strictly increasing functions. This property becomes useful when non linear effects occur.

Another measure $\chi(I_1, I_2)$ less expensive is defined as :

$$\chi(I_1, I_2) = 1 - \frac{2 d_m^{[\frac{n}{2}]}}{[\frac{n}{2}]}$$

We show the insensibility of these two measures to random noise and to rank distortion which can occur due to specular reflection and discontinuity.

Consider the following example of a reference window of dimension 3×3 with the intensities I_1:

$$\begin{array}{ccc} & M & \\ 10 & 30 & 70 \\ 20 & 50 & 80 \\ 40 & 60 & 100 \end{array}$$

Its rank matrix π_1 is:

$$\begin{array}{ccc} 1 & 3 & 7 \\ 2 & 5 & 8 \\ 4 & 6 & 9 \end{array}$$

If another window in the second image is:

$$\begin{array}{ccc} 10 & 30 & 70 \\ 20 & 50 & 80 \\ 40 & 60 & A \end{array}$$

If A is between 81 and 255, π_2 is always :

$$\begin{array}{ccc} 1 & 3 & 7 \\ 2 & 5 & 8 \\ 4 & 6 & 9 \end{array}$$

Which equals to π_1, and $\kappa = 1$.

If A takes the value 75, then κ decreases to 0.8. If A takes a value between 0 and 10, then π_2 becomes

$$\begin{array}{ccc} 2 & 4 & 8 \\ 3 & 6 & 9 \\ 5 & 7 & 1 \end{array}$$

However, the value of κ remains 0.8. This is in sharp contrast to the Kendall's τ and Spearman ρ, which fall steeply to 0.556 and 0.4, respectively. If pixel A takes a value 0, the the linear correlation NCC drops to 0.311.

Similar properties can be shown for χ. In conclusion, these two measures are robust to specularly reflexion and discontinuities .

Non parametric correlations can deal with non linear effects of intensity. They do not eliminate outliers, but use some more stable measures.

In case of rotation, pixel-to-pixel correspondence is not given and it does not seem straight to eliminate the outliers. Instead, it seems more possible to use some non parametric invariants, which are less sensitive to outliers. Let's see it in the next section.

2 Invariant method

In this section, we shall present the invariant method. Firstly we present some generalities to invariant followed by the differential method suggested by Schmid [25]. After an analysis of its drawbacks, we propose a novel non parametric method and its application to image matching problem.

2.1 Introduction to invariants

In this section, we give a short introduction of invariants [22, 9, 27, 25, 30].

Here we are interested in intensity invariants, [22] provides a good introduction to geometric invariants.

Let I_1 and I_2 be two images of a same scene taken under different environments (camera with other parameters, motion of the camera, illumination change, etc). Invariants are properties that remain unchanged under a given set of admissible changes. Two types of change are considered:

1. Change in intensity, i.e. $I_2(x, y) = f(I_1(x, y))$.

This occurs when the camera remains fixed and only illumination changes. f is considered as a increasing function. It can usually be locally approximated by a translational or an affine function, i.e.

$$I_2(x,y) = I_1(x,y) + \mu \text{ or } I_2(x,y) = \lambda I_1(x,y) + \mu.$$

2. Change in geometry, i.e. $I_2(x,y) = I_1(g(x,y))$, where g is a mapping from the two-dimensional image plane to itself.

   The simplest cases for g are translations, which can also be treated by correlation, Hausdorff distance and frequential methods. More generally, in planar patch case, it is well know that g is a homography. If the stereo patch is small enough, then g can be locally approximated by an affine function simulating a parallel projection [29].

Some work can be found in [9], where color distribution was used as invariant primitives, [27] developed a general framework on constructing invariants by averaging method, [30] used moment to build invariants. While these authors use global invariant, [25] presents a local differential invariant method.

In absence of global structure, local methods should be used and we introduce more about [25] in the next section.

2.2 Differential invariant

A characterization of the intensity has been proposed by Koenderink [15] and implemented by Schmid [25], by using differential invariants under the group of similitude between the images. This characterization is based on functions of derivatives invariant to image rigid motion.

It is well know the the calculation of the derivatives is ill-conditioned, it is necessary to calculate the derivatives on smoothed data to struggle against noise. A common choice is to use the Gaussian function as the convolution kernel.

Having calculated the derivatives of a function on a point up to N^{th} order ($N = 3$ has been used), differential invariants under the group of rigid motion can be calculated.

Differential invariants have been successfully used for matching and recognition. After having extracted the interest points [10] in the images, these points are matched by comparing the Mahalanobis distance between the two corresponding invariant vectors. The recognition is achieved using the voting technique, hashing technique and semi-local constraints [26].

Differential invariant method has however the following drawbacks :

- Noise.

  The calculation of derivatives is unstable against noise, smoothing the signal improves the stability, but also damages useful information.

- Non linear effect.

  In case of illumination change, some non linear effect can occur. As indicated in [5], when different cameras are used for stereo and have different responses to image irradiance, each sensor output I is related to image irradiance E as $I = gE^{\frac{1}{\gamma}} + m$, where g is the camera gain, m is the reference bias factor, and γ accounts for image contrast. Let the image surface be Lambertian, i.e. the image irradiance from any point is identical for both sensors. In the case where γ is constant, there is a linear relation between two sensor outputs I_1 and I_2. In general, γ is not constant and the linearity between the sensor outputs is lost. The differential invariants can not deal with such phenomena.

- Partial occlusion.

  In the case where the condition $I_2(x,y) = f(I_1(g(x,y)))$ (f is an affine function and g is a rigid motion) is only verified be some pixels, the differential invariant method fails. This can occur at the disparity discontinuity, specular reflexion, etc.

  In translational case, this phenomena can be dealt with partial or non parametric correlation methods (section 1.2), while in general case, pixel-to-pixel correspondence is not given and it seems not straightful to eliminate the outliers. Non parametrical methods seem more possible, as we shall see in the next section.

2.3 Non Parametric invariant

In this section, we shall present a non parametric rotational invariant method. Polar coordinate system is used, in which the rotations have a simpler form like

$$\begin{cases} r' &= r \\ \theta' &= \theta + c. \end{cases}$$

2.3.1 Obtaining non parametric relations

Supposing only the orders between pixels remain unchanged, that is to say: if $I_1(r_1, \theta) > I_1(r_2, \theta_2)$, then $I_2(r_1, \theta_1') > I_2(r_2, \theta_2')$, where I_1 and I_2 are two corresponding window by a rotation of degree c, and $\theta_i' = \theta_i + c$ $(i = 1, 2)$.

From an intensity window I, we define a comparison function

$$C_I(r_1, \theta, r_2, \theta + \Delta_1) = 1_{I(r_1,\theta) > I(r_2, \theta+\Delta_1)} = \begin{cases} 1 & I(r_1, \theta) > I(r_2, \theta + \Delta_1), \\ 0 & \text{otherwise.} \end{cases}$$

Let us see an example.

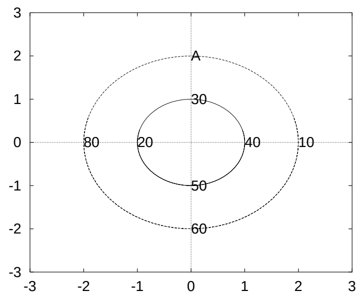


Figure 3: Non parameterization: original image (I_1).

In the figure 3, if $r_1 = r_2 = 2$, $\theta = 0, \Delta_1 = \frac{\pi}{2}$, then

$$C_{I_1}(r_1, \theta, r_2, \theta + \Delta_1) = \begin{cases} 1 & \text{if } A < 10, \\ 0 & \text{otherwise.} \end{cases}$$

If A is perturbeted by artifacts then the influence to $I_1(r_1, \theta, r_2, \theta + \Delta_1)$ is limited to a change of at most 1 even if the perturbation is large.

If the second image window I_2 is the first one rotated by c, the $C_{I_2}(r_1, \theta + c, r_2, \theta + c + \Delta_1) = C_{I_1}(r_1, \theta, r_2, \theta + \Delta_1)$.

As shown in the figure 4, the two image windows I_1 and I_2 correspond to each other by a rotation of degree $\frac{\pi}{2}$ and a non linear effect f on luminance. Suppose f is increasing, then $C_{I_2}(r_1, \theta + \frac{\pi}{2}, r_2, \theta + \frac{\pi}{2} + \Delta_1) = C_{I_1}(r_1, \theta, r_2, \theta + \Delta_1)$.

2.3.2 Eliminating rotation effect by integration

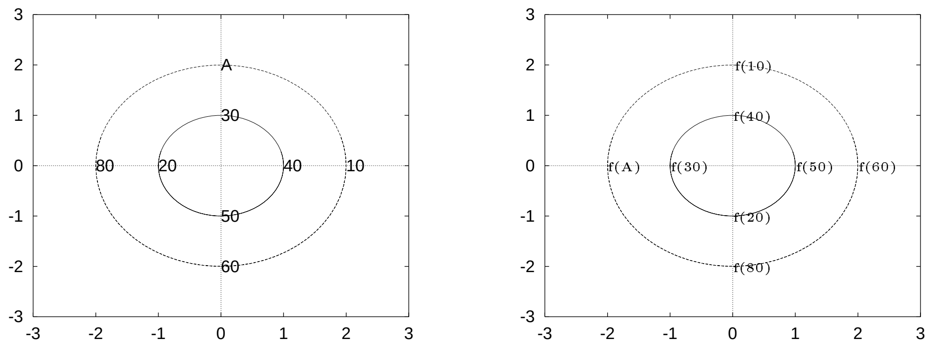
For any r_3, r_4, Δ_2 and Δ_3, we have, $C_{I_2}(r_1, \theta + c, r_2, \theta + c + \Delta_1) - C_{I_2}(r_3, \theta + c + \Delta_2, r_4, \theta + c + \Delta_3) = C_{I_1}(r_1, \theta, r_2, \theta + \Delta_1) - C_{I_1}(r_3, \theta + \Delta_2, r_4, \theta + \Delta_3)$.

Taking the absolute integration to eliminate the rotation degree c, that is to say, we define

$$N_I(r_1, r_2, r_3, r_4, \Delta_1, \Delta_2, \Delta_3) = \frac{1}{2\pi} \int_0^{2\pi} |C_I(r_1, \theta, r_2, \theta + \Delta_1) - C_I(r_3, \theta + \Delta_2, r_4, \theta + \Delta_3)| \, d\theta.$$

From definition N_I is invariant to rotation and increasing grey level transforms.

Integrating on the group of transformation for building invariant is a classical method nicely studied by Schulz-Mirbach [27] for vision application. Here we are in the simplest case: the rotation group is compact, integration is then directly applicable.



(I_1) (I_2)

Figure 4: Image pair under rotation and intensity change.

In our experiments, we have chosen $r_1 = r_2 = s$, $r_3 = r_4 = t$ and $\Delta_3 = \Delta_1 + \Delta_2$. Then we have

$$N_I(s,s,t,t,\Delta_1,\Delta_2,\Delta_1+\Delta_2) = \frac{1}{2\pi}\int_0^{2\pi} |C_I(s,\theta,s,\theta+\Delta_1) - C_I(t,\theta+\Delta_2,t,\theta+\Delta_2+\Delta_1)|\, d\theta.$$

So for each r, several $I(r,\theta)$ are estimated, using an interpolation method. $C_I(r,\theta,r,\theta+\Delta_1)$ are computed for different Δ_1, and the integration over θ can be done to eliminate the rotation effect.

This invariant has the following properties :

1. it is stable to non linear grey level transforms, since non parametric grey level relations are used;
2. it is less sensitive to outliers, as shown in the translational case, huge values have limited changes to this invariant;
3. and therefore it is less sensitive to noise than parametric invariants, since only the orders of the intensity pairs are considered.

After having got the invariant vectors, the L_1 norm is used to measure their distance, we are led to minimizing this norm among several candidates to find the best one.

Let's see the these properties by the following illustrating example (figure of $(r,\theta) \rightarrow C_{I_1}(r,\theta,r,\theta+\frac{\pi}{2}) = \begin{cases} 1 & \text{if } I_1(r_1,\theta) > I_1(r,\theta+\Delta) \\ 0 & \text{otherwise.} \end{cases}$)

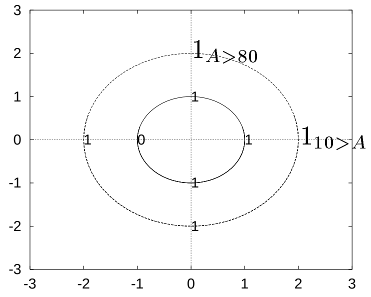
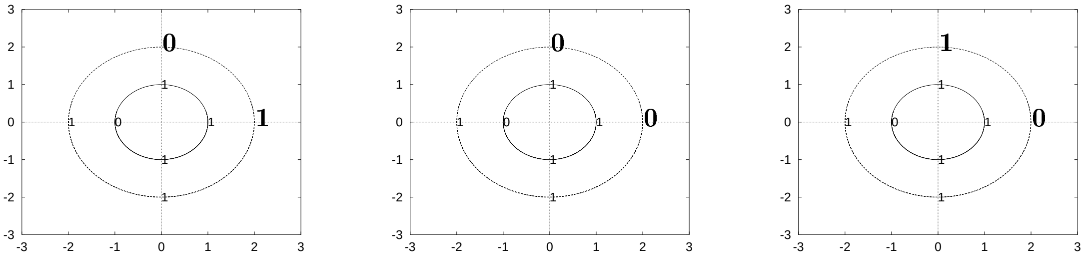


Figure 5: Non parameterization: transformed image of I_1 (Figure 3).

Figure 4 shows the rotation effect is eliminated by the integration: $N_{I_1}(2,2,1,1,\frac{\pi}{2},\frac{\pi}{2},\pi) = N_{I_2}(2,2,1,1,\frac{\pi}{2},\frac{\pi}{2},\pi)$.

In figure 6, for (a), this value is 0; for (b), this value is 0.25; for (c), this value is 0.5. It is less sensible to parametric methods. And it is invariant to rotation due to the integration around the circle.



(a: $A < 10$) (b: $10 < A \leq 80$) (c: $80 < A$)

Figure 6: Non parameterization: transformed image in different cases.

2.3.3 Fast pattern rejection

The invariant vector used has four parameters s, t, Δ_1 and Δ_2, and its dimension can be large. For example, in our experiments, s and t range among 15 values ($\{1..15\}$), Δ_1 and Δ_2 varie among 6 values ($\{1..6\}$), and its dimension is $15 \times 15 \times 6 \times 6 = 8100$. It will be very expensive to compute all the distances between each pair of invariant vectors.

However, given a request invariant vector and the invariant vector base, almost all invariant vectors in the base do not resemble the request one, except several ones. So we lead to find some simple tests to quickly reject the unlike vectors and only few distances are computed for the comparison.

The first test used is the "brightness". For each image window, its center has a relative brightness in the window, which can be defined as the percentage of pixels in the window darker than the center.

This measure is obviously invariant to rotation and non linear (but increasing) illumination change. It is also stable to noise and robust to outliers. If two image windows correspond by a rotation and an illumination change, even if some parts do not correspond, this measure should not differ two much.

In our experiments, a threshold of 0.2 has been chosen for the pattern rejection. If two windows have brightnesses which differ more than 0.2, they are considered as not to be matched.

The second test relies on a measure called "total change". If two vectors N_{I_1} and N_{I_2} are close to each other, then the sums of their components should not differ too much. We split the index set into several subsets and computed several sums over these subsets for the test. This allows us to reject more unlike patterns.

More formally, if $N_{I_1}[i]$ and $N_{I_2}[i]$ close to each other, then $\sum_{i \in Ind'} N_{I_1}[i]$ and $\sum_{i \in Ind'} N_{I_2}[i]$ should not be too large, for any subset Ind' of the total index set Ind.

Suppose $0 \leq N_{I_1}[i], N_{I_2}[i] \leq M$, the difference $\left|\sum_{i \in Ind'} (N_{I_1}[i] - N_{I_2}[i])\right|$ divided by $M\ card(Ind')$ for normalization.

For a specified threshold t_1, if the normalized difference between the two sums associated to two vectors $\left|\sum_{i \in Ind'} (N_{I_1}(i) - N_{I_2}(i))\right| / (M\ card(Ind'))$ is greater than t_1, the two patterns are considered as not corresponding to each other. In our experiments, 0.1. has been taken as as the value of t_1.

The third test uses sub-patterns, If $N_{I_1}[i]$ and $N_{I_2}[i]$ are close to each other and Ind' is a random subset of the total index set Ind, then $\sum_{i \in Ind'} |N_{I_1}[i] - N_{I_2}[i]|$ is usually not too big.

For a specified threshold t_2, if $\sum_{i \in Ind'} |N_{I_1}[i] - N_{I_2}[i]| / (M\ card(Ind')) > t_2$, N_{I_1} and N_{I_2} are considered not alike. Since we choose $|Ind'|$ much smaller than $|Ind|$, we can quickly reject many false candidates. In our experiments, t_2 has been chosen to be 0.1.

The three tests cited above allow us to reject many false candidates and make the large set of non invariant usable in practice.

2.4 Steps of the matching process

We now describe the different stages of our matching algorithm. At first interest points are extracted. Obviously, the extraction quality influences the matching quality. The detector used should be invariant to rotations, repeatable in position. The Harris detector [10] has been chosen as it has the best repeatability.

After the extraction, these points are characterized by the N_I as described in section 2.3.2. For each pair of vectors, a preselection by the three tests proposed in section 2.3.3 is done to reject the false matches. The norm L_1 is then used to compute the distances between the remained pairs after the preselection, and the best one is decided to be that minimizing the norm L_1. Cross-matching can also been used to improve the matching-quality.

3 Experiments

In this section, we present experimental results. The following cases are considered for pixel-matching:

- Geometric change.
- Intensity change.
- Three-dimensional scene with change of viewing point.

For every case, the comparison with the differential method is shown and the improvement is very significant.

In section 3.1, we present two methods for automatic evaluation of the results; in section 3.2, results with different rotations are displayed; in section 3.3 results under different grey level transforms are shown; in section 3.4 results on three-dimensional scenes are shown.

Another criterion is the percentage of total matches in relation to the number of points extracted. This percentage varies depending on the image as well as on the type of transformation between the images; the average percentage of matched points is greater than 50%.

The parameters for the non parametric methods are chosen as described in section 2.3.3, the parameters for the differential methods [25] are $\sigma = 3$ and $r = 15$. Comparisons using other parameters are also done (for exemple, using a circle of radius 10), and in general, the non parametric method gives the same percentage of total matches as the differential one, but gives a greater ($> 10\%$) percentage of good matches.

3.1 Automatic evaluation of the results

In order to evaluate the matching results, two geometric constraints between two images are used. In case of a planar scene, the correspondence of two images of this scene can be modeled by a homography. This homography defines a pixel to pixel relation between two images. Using this homography, a match can be determined to be correct or not, resulting in the percentage of correct matches.

In our experiments, the homography between two images has been estimated using the least-square median method [21] with the found matches. Thanks to the robustness of the least-square median method, the homography can be correctly estimated if more than half of the matches are correct.

In case of non-planar scenes, the epipolar constraint between two views is used as the evaluation criteria. Each point in the first image corresponds to an epipolar line in the second one. A match (p_1, p_2) is evaluated as correct if the point p_2 lies on the epipolar line corresponding to p_1. Although this evaluation is not perfect since a pixel lying on the corresponding epipolar of another pixel might be its false match, the probability that a mismatch satisfies the epipolar constraint is very low, this evaluation is still a good estimation of the number of correct matches.

In our experiments, the epipolar geometry between two images has been estimated using the least-square method [21] also with the found matches. It is robust to outliers up to 50%.

3.2 Geometric change

Two different planar scenes have been used to test invariance to image rotation. The first one is a painted folder called *Sanja* (Figure 7), the second is the painting of *Vangogh* (Figure 8).

For each scene, a sequence of 40 images have been taken between which the camera has been rotated around the optical axis. The matching result depends on the rotation degree, as shown in Figure 9 (*Sanja*) and 10 (*Vangogh*).

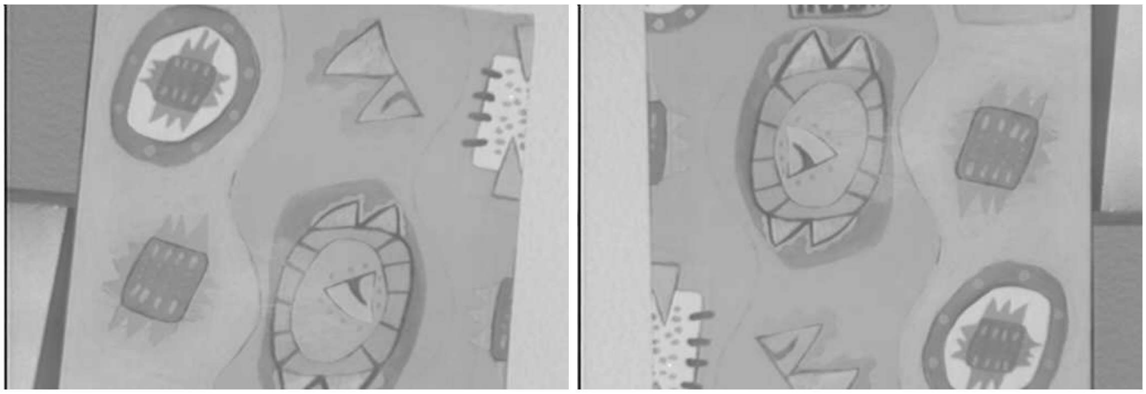


Figure 7: Images *Sanja* under geometric change.

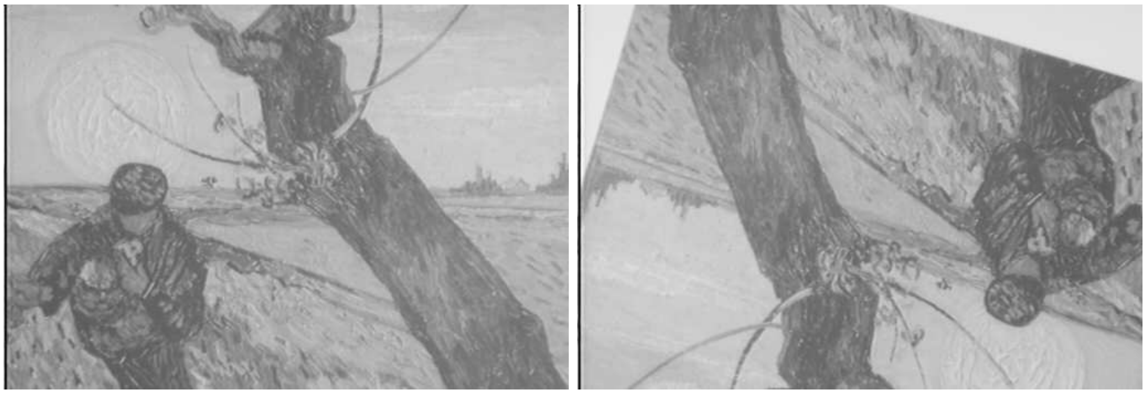


Figure 8: Images Vangogh under geometric change.

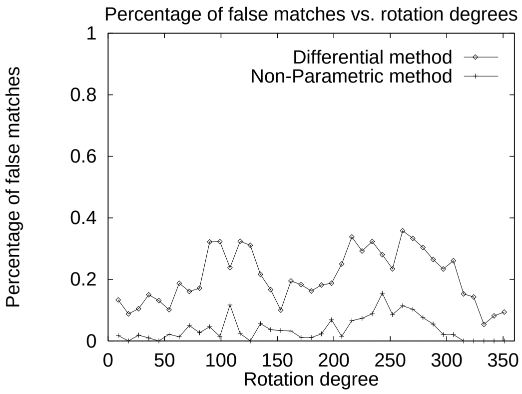


Figure 9: Percentage of false matches vs. the rotation angle for images *Sanja* (Figure 7).

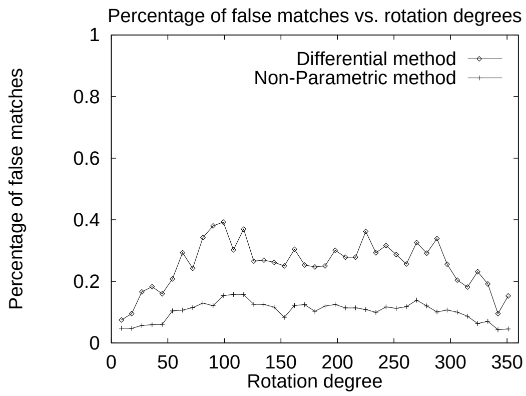


Figure 10: Percentage of false matches vs. the rotation angle for images *Vangogh* (Figure 8).

The improvement comparing to the differential method is significant for two image sequences, especially for images *Vangogh*. Differential method suffers from textures contained in images, due to the bad interest points location, as is the case for images *Vangogh*, whereas non-parametric method is much more stable. Notice also non-parametric gives always fewer than 20% false matches.

3.3 Intensity change

The same two scenes have been use to test invariance to illumination change. The first one *Sanja* is displayed in Figure 11, and the second *Vangogh* is displayed in Figure 12.

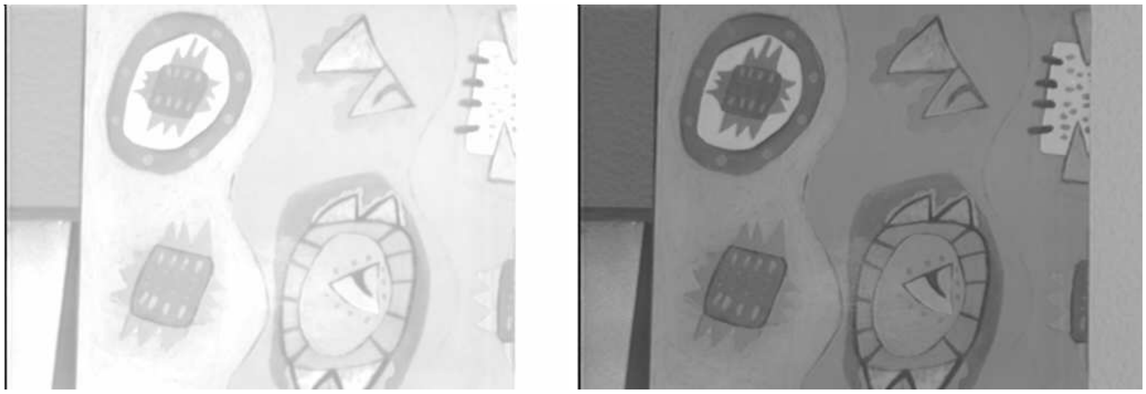


Figure 11: Images *Sanja* under intensity change.

Figure 12: Images *Vangogh* under intensity change.

For each scene, a sequence of 40 images have been taken between which the camera has changed the aperture. The matching result depends on the illumination change, as shown in Figure 13 (*Sanja*) and 14 (*Vangogh*).

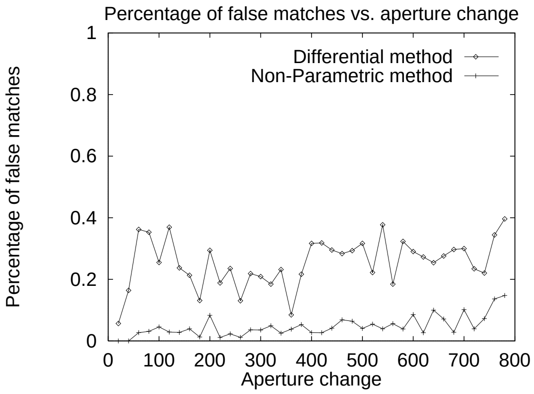


Figure 13: Percentage of false matches vs. the illumination change for images *Sanja* (Figure 11).

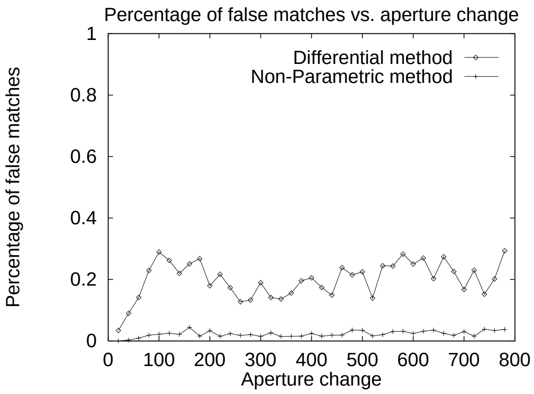


Figure 14: Percentage of false matches vs. the illumination change for images *Vangogh* (Figure 12).

The improvement over the differential method is greater than the case of geometric change. False matches are always fewer than 10%. This occurs from the non linear intensity change, while non parametric method is robust to this kind of change, differential method is not.

3.4 Three-dimensional scene

We present results for three dimensional scenes. The epipolar geometry has been used for the evaluation.

Experiments have been made on outdoor image as shown in figure 15. There are many repetitive patterns in the image. Using the non parametric method, we obtain a score better than 90% (Table 1). This can be considered satisfactory for such a scene. Also the improvement over the differential method is significant.

a b c

Figure 15: Three-dimensional outdoor scene: *tribunal*.

image pair	differential method	non-parametric method
a and b	36.8%	6.4%
a and c	41.8%	9.3%
b and c	25.4%	9.3%

Table 1: False matches on images *tribunal* (figure 15).

Image	Position	Scale	Orientation
Im1	P0	s0	r0
Im2	P0	s1	r0
Im3	P0	s0	r1
Im4	P0	s1	r1
Im5	P1	s0	r0
Im6	P1	s1	r0
Im7	P1	s1	r1
Im8	P1	s2	r1
Im9	P1	s2	r0
Im10	P2	s0	r0
Im11	P2	s0	r1
Im12	P2	s0	r2

Table 2: Different camera conditions for images *Salle*. $s0 = 1.7, s1 = 1.77, s2 = 1.85, r0 = 0$ (degrees), $r1 = 10, r2 = 20$.

Experiments have been made on images shown in figure 16. The transformation between the images consists of a scene rotation, camera rotation, disparity discontinuity, etc.

The images in the sequence have been taken under several camera changes, including position change, scale change, and orientation change, as summarized in the table 2.

Figure 16: Two images of the three-dimensional indoor scene: *salle*.

Figure 18 shows the displacement vectors for the correct matches in figure 16. The global rotation is clearly visible, as are the few wrong matches.

The scale change is small and has little influence to the matching quality. Camera position change brings disparity discontinuities and damages the matching results. Camera orientation change has also some influence on the matching results, but less than than the position change.

Figure 18 shows the matching results between image1 and all other images in the sequences.

As shown in Figure 18, the matching results is very good for images 2, 3, 4 which are taken at the same position as image 1; for images 5 to 9, the results are less good and for images 10 to 12, they are still worse, because of the camera displacement. For every image, the improvement over the differential method is more

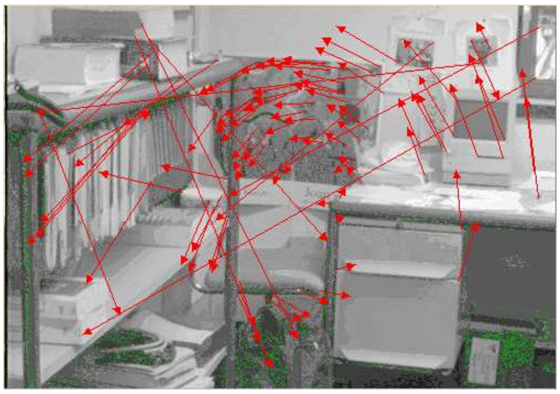


Figure 17: Disparity estimated between images in Figure 16.

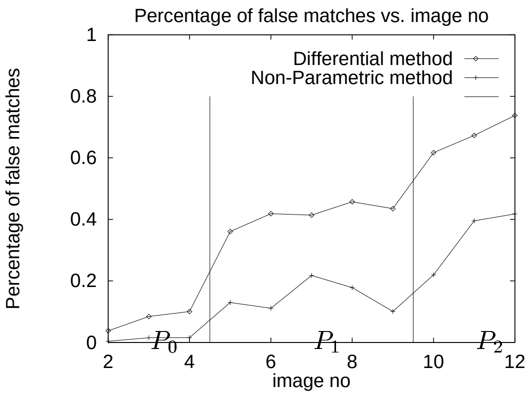


Figure 18: Percentage of false matches vs. the image no for images "Salle" (Figure 16).

significant than over the images of planar scenes, because the differential method is fragile to the disparity discontinuity, while the non-parametric one is much less sensible to it.

4 Sub-pixel matching method

The sub-pixel matching is an interest problem which is useful in photogrammetry, auto-calibration, etc. Previous work of Ackermann [1], Gruen [8], Lan [17] rely on correlations, which have some difficulty in presence of huge rotation.

If the interest point extraction has a sub-pixel precision, the matching of these points will also be precise. However, the interest point extraction often has an error up to 2 pixels, it is therefore not hopeful to have a precise extraction method.

We propose another hierarchic method as following: for a pixel p_1 in the image I_1, if we have its initial match p_2 in the image I_2, we can choose some neighbors of p_2: $p_{21}, p_{22}, p_{23}, p_{24}$, compare them with p_1, and choose the best one as the new p_2.

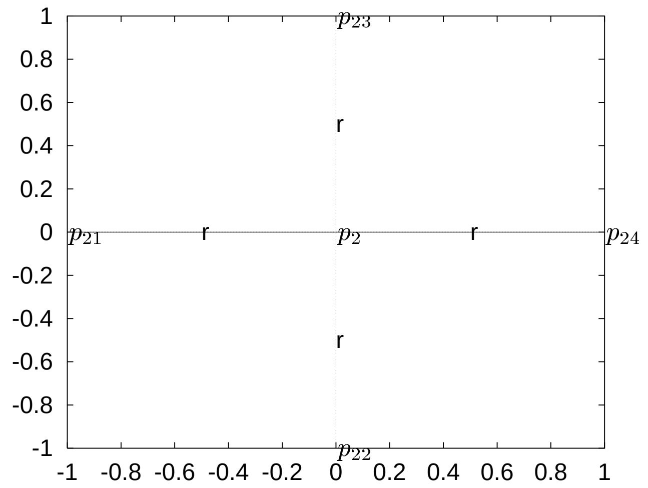


Figure 19: Sub-pixel matching

This process is repeated for different displacement of size r, the initial value of r has been set to 1, and each time, it is divided by 2. Usually the iteration number is about 5.

The advantage of this simple algorithm comparing to the "robust" method presented in [17] is it relies on only one initial match, and the dense disparity map is not required; its drawback is that it can not eliminate the false matches in the initialization and can not deal with affine transforms. However, in large rotation case, it seems difficult to have get dense disparity map, and this method seems feasible to get precise matches.

5 Experimental results on sub-pixel matching

Sub-pixel matching have been tested on images *Sanja* and *Vangogh* using the homography between images and also on images of non planar scene *Salle* and *Tribunal* using the epipolar geometry (section 3.1).

The same strategy has also been tested using the differential invariants. Unfortunately, the precision we got is at level 0.4 pixels, which is much worse than the non parametric method. The reason is perhaps that the differential invariants are too sensible to the interpolation, which is used in the hierarchic search. So their results are not shown in the following graphs.

5.1 Geometric change

Two different planar scenes have been use to test invariance to image rotation. The first one is a painted folder called *Sanja* (Figure 7), the second is the painting of *Vangogh* (Figure 8).

For each scene, a sequence of 40 images have been taken between which the camera has been rotated around the optical axis. The matching result depends on the rotation degree, as shown in Figure 20 (*Sanja*) and 21 (*Vangogh*).

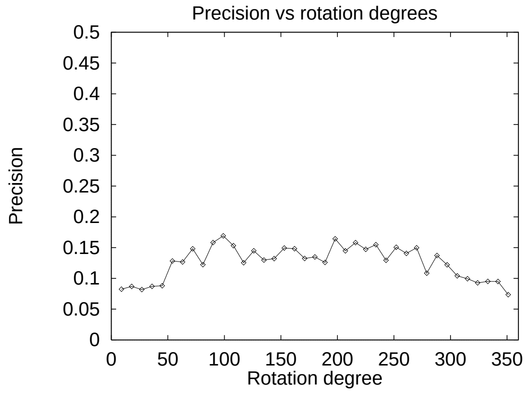


Figure 20: Matching precision vs. the rotation angle for images *Sanja* (Figure 7).

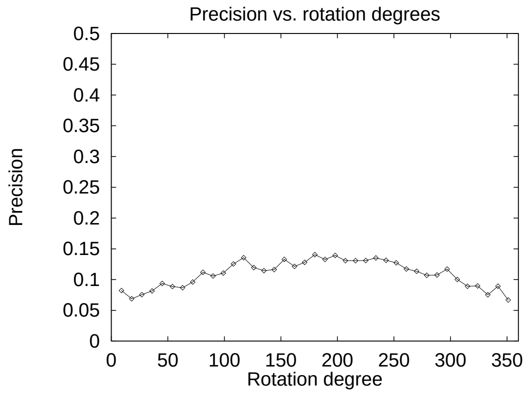


Figure 21: Matching precision vs. the rotation angle for images *Vangogh* (Figure 8).

The precision we get is usually between 0.1 and 0.2 pixels, sometimes better, such values are good when considering these difficult conditions.

5.2 Intensity change

The same two scenes have been use to test invariance to illumination change. The first one *Sanja* is displayed in Figure 11, and the second *Vangogh* is displayed in Figure 12.

For each scene, a sequence of 40 images have been taken between which the camera has changed the aperture. The matching result depends on the illumination change, as shown in Figure 22 (*Sanja*) and 23 (*Vangogh*).

The precision we get is between 0.1 and 0.2 pixels, which is good enough under these difficult cases.

5.3 Three-dimensional scene

We present results for three dimensional scenes. The epipolar geometry has been used for the evaluation.

Experiments have been made on outdoor image as shown in figure 15. There are many repetitive patterns in the image. Using the non parametric method, we obtain a precision of about 0.2 pixels (Table 5.3). This can be considered satisfactory for such a scene.

Experiments have been made on images shown in figure 16. The transformation between the images consists of a scene rotation, camera rotation, disparity discontinuity, etc.

The images in the sequence have been taken under several camera changes, including position change, scale change, and orientation change, as summarized in the table 2.

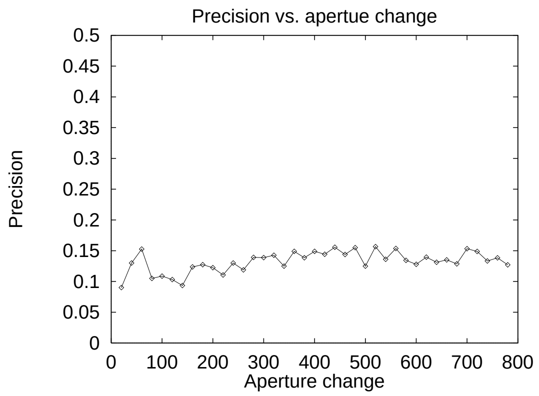


Figure 22: Matching precision vs. the illumination change for images *Sanja* (Figure 11).

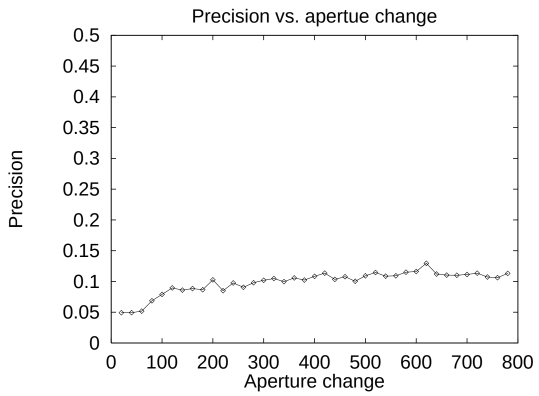


Figure 23: Matching precision vs. the illumination change for images *Vangogh* (Figure 12).

Image pair	Precision (pixels)
a and b	0.20
a and c	0.23
b and c	0.12

Table 3: Precision on images *Tribunal* (Figure 15) using the epipolar geometry.

Figure 24 shows the matching results between image1 and all other images in the sequences.

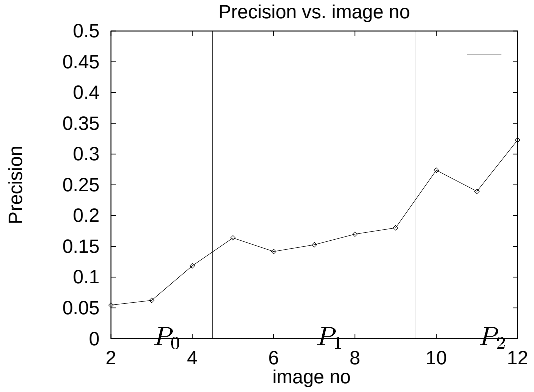


Figure 24: Precision vs. the image no for images *Salle* (Figure 16).

6 Conclusion

Non-parametric invariants are used to characterize points in the image. As they are less sensitive to outliers, they are well suited in case of partial occlusion. They are also less sensitive to noise and non linear intensity changes. Experimental results on many images are shown and they validate the approach. Improvement comparing to the differential invariant method is clear, under rotation, intensity change and more in partial occlusion case. At last, a sub-pixel matching algorithm is provided using these invariants; it allows 1/10 pixels accuracy in presence of rotation; this fast and easy algorithm might be useful for the area as photogrammetry, auto-calibration, etc.

References

[1] F. Ackermann. Digital image correlation: performance and potential application in photogrammetry. *Photogrammetric Record*, 64(11):429–439, October 1984.

[2] 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.

[3] N. Ayache. *Stereovision and Sensor Fusion*. MIT-Press, 1990.

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

[5] 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.

[6] D.J. Fleet, A.D. Jepson, and M.R.M. Jenkin. Phase-based disparity measurement. *Computer Vision, Graphics and Image Processing*, 53(2):198–208, March 1991.

[7] W.E.L. Grimson. *From Images to Surfaces. A Computational Study of the Human Early Visual System.* The MIT Press, Cambridge, MA, USA, 1981.

[8] A.W. Gruen. Adaptative least squares correlation: a powerful image matching technique. *S. Afr. Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3):175–187, 1985.

[9] G. Healey and D. Slater. Global color constancy: Recognition of objects by use of illumination-invariant properties of color distributions. *Journal of the Optical Society of America A*, 11(11):3003–3010, 1994.

[10] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kuebler. Simulation of neural contour mechanism: from simple to end-stopped cells. *Vision Research*, 32(5):963–981, 1992.

[11] R. Horaud and Th. Skordas. Stereo correspondence through feature grouping and maximal cliques. IEEE *Transactions on PAMI*, 11(11):1168–1180, 1989.

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

[13] X. Hu and N. Ahuja. Feature extraction and matching as signal detection. *International Journal of Pattern Recognition and Artificial Intelligence*, 8(6):1343–1379, 1994.

[14] D.P. Huttenlocher and W.J. Rucklidge. A multi-resolution technique for comparing images using the Hausdorff distance. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 705–706, 1993.

[15] J. J. Konderink and A. J. van Doorn. Representation of local geometry in the visual system. *Biological Cybernetics*, 55:367–375, 1987.

[16] 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.

[17] Z.D. Lan and R. Mohr. Direct linear sub-pixel correlation by incorporating neighbour pixels' information: a robust and precise matching method, June 1996. Submitted to MVA.

[18] Z.D. Lan and R. Mohr. Robust location based partial correlation. Inria Rhone Alpes, January 1997. accepted by CAIP97.

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

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

[21] P. Meer, D. Mintz, A. Rosenfeld, and D.Y. Kim. Robust regression methods for computer vision: a review. *International Journal of Computer Vision*, 6(1):59–70, 1991.

[22] J.L. Mundy and A. Zisserman, editors. *Geometric Invariance in Computer Vision*. The MIT Press, Cambridge, MA, USA, 1992.

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

[24] W.J. Rucklidge. Locating objects using the Hausdorff distance. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 457–464, 1995.

[25] C. Schmid and R. Mohr. Matching by local invariants. Technical report, INRIA, August 1995.

[26] C. Schmid and R. Mohr. Combining greyvalue invariants with local constraints for object recognition. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*, June 1996.

[27] H. Schulz-Mirbach. Constructing invariant features by averaging techniques. In *Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel*, pages 387–390, 1994.

[28] H. Shekarforoush, M. Berthod, and J. Zerubia. Subpixel image registration by estimating the polyphase decomposition of cross power spectrum. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*, pages 532–537, June 1996.

[29] S. Ullman and R. Basri. Recognition by linear combinations of models. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 13(10):992–1006, 1991.

[30] L. van Gool, T. Moons, and D. Ungureanu. Affine / photometric invariants for planar intensity patterns. In *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, pages 642–651, 1996.

[31] X. Wu and B. Bhanu. Gabor wavelets for 3D object recognition. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 537–542, 1995.

[32] 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.

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