# CONTRAST INVARIANT REGISTRATION OF IMAGES

*Pascal Monasse*

CMLA, Ecole Normale Supérieure de Cachan
61, Avenue du Président Wilson
94235 Cachan Cedex, France
`monasse@cmla.ens-cachan.fr`

**ABSTRACT**

We propose a method for image registration which seems to be useful under the three following conditions. First, both images are globally and roughly the result of a translation and rotation. Second, some occlusions due to moving objects occur from image 1 to image 2. Third, because of changes of illumination, contrast may have changed globally and even locally. Under such unfavorable conditions, correlation-based global registration may become inaccurate, because of the global compromise it yields between several displacements. Our method avoids these difficulties by defining a set of local contrast invariant features in order to achieve contrast invariant matching. A voting procedure allows to eliminate ”wrong” matching features due to the displacement of small objects and yields sub-pixel accuracy. This method was tested successfully for registration of watches with moving hands and for road control applications.

## 1. INTRODUCTION

There are roughly three kinds of techniques for registration. We refer to the excellent review by [2] and references therein. Most techniques are based on global or local correlation and lead to fast computations in the Fourier domain [5], [13, chapter 9]. A second kind of method (like [11]) first computes some image salient features, which should be sufficiently stable to appear in both images. These features may be semi-local (regions coming from a segmentation procedure) or very local, like edge points and corners. They may be very heavy in computational cost, since local features may have lots of matches. The method we have explored can be classified in semi-local methods. Its main idea is to match grey-level based neigborhoods, that is, connected regions whose grey level is between two values. Let us call these regions sections. If we change the contrast of an image, the set of all sections is globally invariant.

This is a slight generalization of a classical mathematical morphology representation of the image by connected components of level sets [6, 9]. Also, the idea of defining grey-level based regions in the image is proposed by Yaroslavsky [13] in order to define adaptive filtering and in [12] where it is explicitly aimed at stereo-matching. The considered neighborhoods in [12] are not contrast invariant, however. Another related method is [1], where the same neighborhoods are used for “image intersection” in satellite imaging.

In contrast to local correlation methods, the proposed method does not fix any *a priori* neighborhood for matching. This may be an explanation for its better performance as it is for the Yaroslavsky neighborhood based filtering. The semi-global methods permit to take advantage of global enough features for accurate displacement estimation. Also, they are local enough (since there are sections of any size) to raise the hope that many sections in both images are not altered by any occlusion. As we shall see in the experiments, when the images are small-sized and with occlusions, correlation may fail where contrast invariant voting works reliably.

## 2. A CONTRAST INVARIANT APPROACH

### 2.1. The Shapes

We model a change of contrast as a strictly increasing, continuous real function $g$. If $u$ is a first image and $v$ the result of another snapshot of the same scene, then a rough model for the alteration of $u$ into $v$ is $v = g(u)$ [3].

Let us define formally upper and lower level sets of the image by

$$\mathcal{X}^{\mu} u = \{ \boldsymbol{x} \;/\; u(\boldsymbol{x}) \;\le\; \mu \;\} \,,\; \mathcal{X}_{\lambda} u = \{ \boldsymbol{x} \;/\; u(\boldsymbol{x}) \;\ge\; \lambda \;\}$$

and the sections by $\mathcal{X}^{\mu}_{\lambda} = \mathcal{X}_{\lambda} \cap \mathcal{X}^{\mu}$. Then it is easily seen that $\mathcal{X}^{g(\mu)} g \circ u = \mathcal{X}^{\mu} u$ and $\mathcal{X}_{g(\lambda)} g \circ u = \mathcal{X}_{\lambda}$, so that these sets of features are globally contrast invariant (every section of $u$ is a section of $v$ and conversely). Since we checked that the matching algorithm works much faster and reliably enough with connected components of upper and lower level sets, we shall call them “shapes” in the following.

### 2.2. Matching of Components of Level Sets

In order to perform fast matching, we proceed, as it is classical (also used in [1]) by computing global characteristics for each shape: its area, its barycenter (first order moments), and second order moments. These moments can be combined to yield *rotation and translation invariant* characteristics [8]. Consider the matrix of inertia of the shape $S$:

$$I_S = \begin{pmatrix} \mu_S^{2,0} & \mu_S^{1,1} \\ \mu_S^{1,1} & \mu_S^{0,2} \end{pmatrix}$$

where $\mu_S^{(i,j)} = \int_S (x \Leftrightarrow x_S)^i (y \Leftrightarrow y_S)^j dx dy$ and $(x_S, y_S)$ is the barycenter of $S$. Now consider the rotated and translated version of $S$, $R_\theta S + \boldsymbol{t}$. Then $I_{R_\theta S + \boldsymbol{t}} = R_\theta \; I_S \; R_{-\theta}$ so that $Det \; I_{R_\theta S + \boldsymbol{t}} = Det \; I_S$ and also $Tr \; I_{R_\theta S + \boldsymbol{t}} = Tr \; I_S$. Two shapes $S$ and $S'$ extracted from $u_1$ and $u_2$ respectively, are considered to be candidates to matching if:

$$S \to S' \Leftrightarrow \left\{ \begin{array}{ccc} A_S & \cong & A_{S'} \\ Det\ I_S & \cong & Det\ I_{S'} \\ Tr\ I_S & \cong & Tr\ I_{S'} \end{array} \right.$$

Notice that in order to allow small deformations of the shapes and to avoid quantization effects, we do not impose a strict equality. A typical difference of a characteristic can be between 10 % and 20 %.

### 2.3. Voting

The voting procedure we use is similar to the one described in [4]. We compute our registration as the isometry of the plane compatible with a maximum number of correspondences $\boldsymbol{x}_i \to \boldsymbol{y}_j$ (where $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ are the barycenters of $S_i$ and $S'_j$) and $S_i$ in the first image and $S'_j$ in the second image match. Now, there are infinitely many rotation-translations that map $\boldsymbol{x}_i$ to $\boldsymbol{y}_j$ for a given $(i,j)$. If, in addition, we can find another correspondence $S_k \to S'_l$ such that $\|\boldsymbol{x}_i \Leftrightarrow \boldsymbol{x}_k\| = \|\boldsymbol{y}_j \Leftrightarrow \boldsymbol{y}_k\| \neq 0$, then there is only one rotation-translation compatible with both correspondences:

$$\begin{aligned} \sin\theta &= \frac{\left\|(\boldsymbol{x}_i \Leftrightarrow \boldsymbol{x}_k) \wedge (\boldsymbol{y}_j \Leftrightarrow \boldsymbol{y}_k)\right\|}{\|\boldsymbol{x}_i \Leftrightarrow \boldsymbol{x}_k\|^2} \\ \cos\theta &= \frac{(\boldsymbol{x}_i \Leftrightarrow \boldsymbol{x}_k).(\boldsymbol{y}_j \Leftrightarrow \boldsymbol{y}_k)}{\|\boldsymbol{x}_i \Leftrightarrow \boldsymbol{x}_k\|^2} \\ \boldsymbol{t} &= \boldsymbol{y}_j \Leftrightarrow R_\theta \boldsymbol{x}_i (= \boldsymbol{y}_l \Leftrightarrow R_\theta \boldsymbol{x}_k) \end{aligned}$$

The voting procedure accepts this rotation-translation (that is in a 3-D parameters space). The final selected rotation-translation is the one getting the maximum of votes.

## 3. THE SUBPIXEL ACCURACY REGISTRATION

Note that the result obtained is not very precise, due to the necessary quantization of the unknown parameters, and even refining the quantization step is not very satisfactory because the peak of votes would simply become blurred. Typically, this vote gives a precision of the order of one pixel. If we want a better estimation (and in most applications it is very important [10]), we begin by selecting all shapes in both images which have obtained a correspondence compatible with the winning rotation-translation. We say that $S_i$ and $S'_j$ are in correspondence if:

$$S_i \to S'_j \text{ and } \left\|\boldsymbol{y}_j \Leftrightarrow R_\theta \boldsymbol{x}_i \Leftrightarrow \boldsymbol{t}\right\| \le \epsilon \qquad (1)$$

where $\epsilon$ is a given threshold. Then we can proceed to a "bootstrap" estimation of the rotation translation by a least squares minimization:

$$\arg\min_{\theta, \boldsymbol{t}} \sum_{i,j\ /\ S_i \to S'_j} \left\|R_\theta\ \boldsymbol{x}_i + \boldsymbol{t} \Leftrightarrow \boldsymbol{y}_j\right\|^2$$

A difficulty arises, because the problem is (apparently) nonlinear due to the fact that $\theta$ appears in the form of $\cos\theta$ and $\sin\theta$. We introduce a scale factor $s$, so that the problem becomes linear:

$$\arg\min_{s_1, s_2, \boldsymbol{t}} \sum \left\| \begin{pmatrix} s_1 & \Leftrightarrow s_2 \\ s_2 & s_1 \end{pmatrix} \boldsymbol{x}_i + \boldsymbol{t} \Leftrightarrow \boldsymbol{y}_j \right\|^2 . \qquad (2)$$

If all goes well, the optimal scale parameter $s = \sqrt{s_1^2 + s_2^2}$ should be close to 1, and then $\cos\theta = s_1/s$ and $\sin\theta = s_2/s$.

Noting $S = (s_1\ s_2)^T$ and taking into account the equality

$$\begin{pmatrix} s_1 & \Leftrightarrow s_2 \\ s_2 & s_1 \end{pmatrix} \boldsymbol{x}_i = A^{(i)} S \text{ where } A^{(i)} = \begin{pmatrix} x_i & \Leftrightarrow y_i \\ y_i & x_i \end{pmatrix},$$

we can rewrite equation (2) into

$$\arg\min_{S, \boldsymbol{t}} \sum \left\|A^{(i)} S + \boldsymbol{t} \Leftrightarrow \boldsymbol{y}_j\right\|^2 . \qquad (3)$$

In order to solve equation (3), we compute the partial derivatives relative to the parameters and equate them to 0. This yields the system

$$\left\{ \begin{array}{ccccc} \sum A^{(i)\,T} A^{(i)}\ S & + & \left(\sum A^{(i)}\right)^T \boldsymbol{t} & = & \sum A^{(i)\,T} \boldsymbol{y}_j \\ \sum A^{(i)}\ S & + & N\,\boldsymbol{t} & = & \sum \boldsymbol{y}_j \end{array} \right.$$

with $N$ being the number of matchings compatible with the modes. After some easy algebraic manipulations, we get

$$\left\{ \begin{array}{lc} \left[\sum A^{(i)\,T} A^{(i)} \Leftrightarrow \frac{1}{N}\left(\sum A^{(i)}\right)^T \left(\sum A^{(i)}\right)\right] S & = \\ \quad \sum A^{(i)\,T} \boldsymbol{y}_j \Leftrightarrow \frac{1}{N}\left(\sum A^{(i)}\right)^T \left(\sum \boldsymbol{y}_j\right) & \\ \qquad\qquad \boldsymbol{t} & = \\ \quad \frac{1}{N}\left[\sum \boldsymbol{y}_j \Leftrightarrow \sum A^{(i)}\ S\right] & \end{array} \right.$$

which allows to compute the vector $S$ and after the vector $\boldsymbol{t}$.

To compute $S$, we have to invert the $2 \times 2$ matrix

$$\sum_i A^{(i)\,T} A^{(i)} \Leftrightarrow \frac{1}{N} \left(\sum_i A^{(i)}\right)^T \left(\sum_i A^{(i)}\right)$$

which by Cauchy-Schwarz inequality is singular if and only if all the $A^{(i)}$ are proportional, that is all the $\boldsymbol{x}_i$ are aligned, which is a very special case.

## 4. EXPERIMENTS

### 4.1. Comparison with Correlation Based Registration

We have compared our voting procedure with the most classical Fourier-based correlation. If the two images $u_1$ and $u_2$ are related by a simple translation, $u_2(\boldsymbol{x}) = u_1(\boldsymbol{x} \Leftrightarrow \boldsymbol{t})$, then the Fourier Shift theorem gives the following relation between their Fourier transforms [5]:

$$\mathcal{F}u_2(\boldsymbol{\xi}) = e^{-i\boldsymbol{\xi}.\boldsymbol{t}} \mathcal{F}u_1(\boldsymbol{\xi})$$

(where $\mathcal{F}u(\boldsymbol{\xi}) = \int u(\boldsymbol{x}) e^{-i\boldsymbol{x}.\boldsymbol{\xi}} d\boldsymbol{x}$) then the unknown translation vector $\boldsymbol{t}$ can be isolated by the formula:

$$e^{-i\boldsymbol{\xi}.\boldsymbol{t}} = \frac{\overline{\mathcal{F}u_1(\boldsymbol{\xi})} \mathcal{F}u_2(\boldsymbol{\xi})}{|\mathcal{F}u_1(\boldsymbol{\xi})|^2}$$

so that

$$\mathcal{F}^{-1}\left(\frac{\overline{\mathcal{F}u_1(\boldsymbol{\xi})} \mathcal{F}u_1(\boldsymbol{\xi})}{|\mathcal{F}u_2(\boldsymbol{\xi})|^2}\right) = \delta_{\boldsymbol{t}} \qquad (4)$$

which provides a very simple algorithm to estimate the translation vector $\boldsymbol{t}$.

We show in figure 1 an example of the registration of an image of a watch, for which the preceding Fourier registration method does not yield the true rotation-translation. Notice that the hands, whose motion is not compatible with the global rotation-translation, do not disturb the accuracy of the morphological registration obviously because of the voting procedure. Figure 2 shows the projection of the votes in the parameter-space along the three axes ($\theta$, $t_x$, $t_y$). Notice the sharpness of the peaks, showing that the votes are unambiguous.

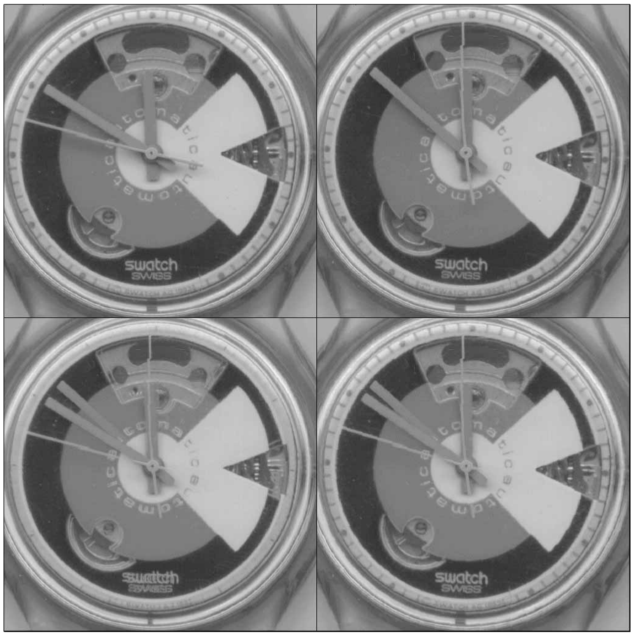

Figure 1: Top: two original images (size:$512 \times 512$) of the same watch at different times (data: SMH Automation). Bottom: the average image of the images after a Fourier registration (left), the average image after contrast invariant registration (right). Notice how the Fourier registration lets all symbols on the watch blur and fade.

Figure 3 is an example of the registration of outdoor images. Due to the wind, the camera rotated between the two shots and the true displacement is more complex than just a rotation-translation. This experiment illustrates the difference between global registration and voting procedures: none of both can yield the "right matching", since this matching is not an isometry. Now the global correlation yields an average displacement, which is true nowhere. The voting procedure selects a part of the image where displacement is accurate (the left hand part of the image) and of course wrong on the other part. Figure 4 shows the votes for the three parameters.

### 4.2. Subpixel Accuracy of the Algorithm

To measure the accuracy of the registration, we took an image of size $1280 \times 1014$ (containing 3250 shapes), applied it a translation and a rotation and then zoomed it out by a factor 2, to obtain an image of size $640 \times 507$. Then a translation of an odd number of pixels in the original image corresponds to half an integer translation in the reduced image. The translation and rotation is made by a *nearest neighbor* interpolation in the original image. The results of the algorithm are shown in Table 1.

From this table, we can observe that the precision concerning the angle is very good. We have a precision of the order of a few hundredths of degree. Concerning the translation, the results seem

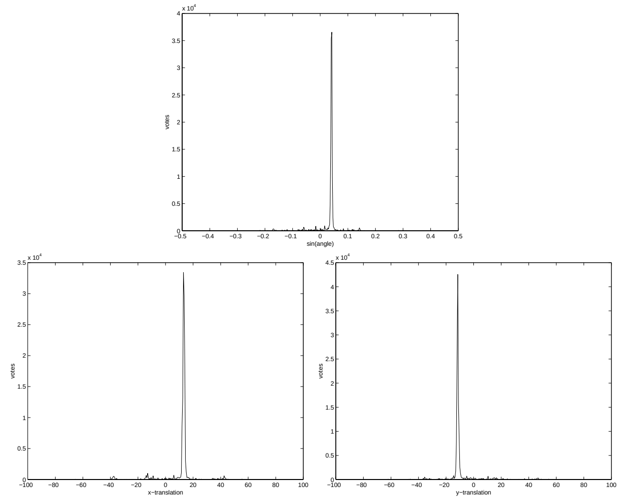

Figure 2: The votes for the three parameters for the images of a watch in figure 1. There are 5072 shapes of area greater than 20 pixels in the first image, and 4961 in the second image. The algorithm found 4065 correspondences.

at first sight less good. Typically, we have a precision of the order of a quarter of pixel, but it is often better. Remember that we used a *nearest neighbor* interpolation to transform the original image, which has typically a precision of half a pixel, which makes a quarter of pixel in the reduced image. So the error of the algorithm is of the same order as the one of the interpolation. We could not hope a better result with this kind of interpolation. Following this idea, we can see that for an integer number of pixels for the translation and no rotation, we have much more correspondences (more than 2500, instead of 1200 or less otherwise) than in the other cases. This corresponds to the case where there is actually no interpolation. In this case, we have a precision of a few thousandths of pixel. In each experiment, the measured scale factor $s$ did not differ from the value $1$ by more than $10^{-6}$.

# Acknowledgments

The author thanks Frédéric Guichard, Jean-Michel Morel and Leonid Yaroslavsky for useful discussions about the subject.

## 5. REFERENCES

[1] Ballester C., Cubero-Castan E., González M., and Morel J.M. "Fast contrast invariant image intersection", *preprint CEREMADE*, 1998.

[2] Brown L.G. "A survey of image registration techniques", *ACM Computing Surveys*, 24(4):325-376, 1992

[3] Caselles V., Lisani J.L., Morel J.M., and Sapiro G. "Shape preserving local histogram modification", *IEEE International Conference on Image Processing*, Lausanne, 1996.

[4] Chang S.H., Cheng F.H., Hsu W.H., and Wu .Z. "Fast algorithm for point pattern-matching: invariant to translations, rotations and scale changes", *Pattern Recognition*, 30(2):311-320, 1997.

Figure 3: Top: two original images (size: $256 \times 256$) of a street scene at different times. Bottom: the average of the images after a Fourier registration (left), the average after a morphological registration (right).

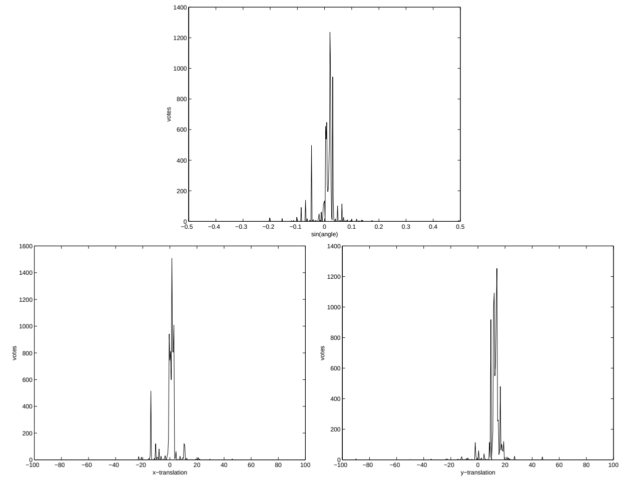


Figure 4: The votes for the three parameters for the outdoor images of figure 3. There are 4979 shapes of area greater than 20 pixels in the first image, and 4612 in the second image. The algorithm found 1589 correspondences.


[5] DeCastro E. and Morandi C. "Registration of translated and rotated images using finite Fourier transforms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):700-703, 1983.

[6] Matheron G. *Random Sets and Integral Geometry*, John Wiley, 1975.

[7] Reddy B.S. and Chatterji B.N. "An FFT-based technique for translation, rotation, and scale-invariant image registration", *IEEE Transactions on Image Processing*, 5(8):1266-1271, 1996.

[8] Reiss T.H. *Recognizing planar objects using invariant image features*, Lecture Notes in Computer Science, 676, Springer Verlag, 1993.

[9] Serra J. *Image analysis and mathematical morphology*, Academic Press, 1988.

[10] Tian Q. and Huhns M.N. "Algorithms for subpixel registration", *Computer Vision, Graphics and Image Processing*, 35:220-233, 1986.

[11] Wang C.Y., Sun H., Yada S., and Rosenfeld A. "Some experiments in relaxation image matching using corner features", *Pattern Recognition*, 17:569-573, 1984.

[12] Wang Y. and Bhattacharya P. "Hierarchical stereo correspondence using features of gray connected components", *IEEE International Conference on Image Processing*, Santa Barbara, 1997, pages 264-267.

[13] Yaroslavsky L. and Eden M. *Fundamentals of digital optics*, Birkhäuser, 1996.


| $\theta, t_x, t_y$ | $\vert\tilde{\theta} \Leftrightarrow \theta\vert$ | $\vert\tilde{t}_x \Leftrightarrow t_x\vert$ | $\vert\tilde{t}_y \Leftrightarrow t_y\vert$ | $Corr.$ |
|---|---|---|---|---|
| $0, 0.5, 0.5$ | $2.5\ 10^{-2}$ | $1.3\ 10^{-1}$ | $1.3\ 10^{-1}$ | 1241 |
| $0, 1, 1$ | $1.2\ 10^{-4}$ | $3.4\ 10^{-4}$ | $3.3\ 10^{-3}$ | 2757 |
| $0, 1.5, 1.5$ | $2.5\ 10^{-2}$ | $1.4\ 10^{-1}$ | $1.6\ 10^{-1}$ | 1194 |
| $0, 2, 2$ | $2.8\ 10^{-4}$ | $1.3\ 10^{-3}$ | $2.7\ 10^{-4}$ | 2741 |
| $0, 10, 10$ | $2.3\ 10^{-4}$ | $9.7\ 10^{-4}$ | $4.0\ 10^{-4}$ | 2652 |
| $0, 10.5, 10.5$ | $1.9\ 10^{-2}$ | $1.2\ 10^{-1}$ | $1.0\ 10^{-1}$ | 1145 |
| $0, 11.5, 11.5$ | $1.7\ 10^{-2}$ | $1.1\ 10^{-1}$ | $1.0\ 10^{-1}$ | 1136 |
| $0.3, 7.5, 1.5$ | $2.4\ 10^{-3}$ | $1.6\ 10^{-1}$ | $2.3\ 10^{-1}$ | 1550 |
| $1, 25, 25$ | $7.0\ 10^{-3}$ | $2.4\ 10^{-2}$ | $1.4\ 10^{-1}$ | 1211 |
| $5, 26.5, 13.5$ | $1.8\ 10^{-3}$ | $3.1\ 10^{-1}$ | $4.2\ 10^{-1}$ | 975 |
| $10, 20, 17.5$ | $2.7\ 10^{-3}$ | $3.6\ 10^{-1}$ | $2.5\ 10^{-1}$ | 883 |
| $20, 30.5, 10$ | $5.7\ 10^{-3}$ | $7.6\ 10^{-2}$ | $1.2\ 10^{-1}$ | 723 |

Table 1: Measurements of the accuracy of the algorithm for several rotations and translations. The exact rotation and translation ($\theta$, $t_x$, $t_y$) is compared to the measured one ($\tilde{\theta}$, $\tilde{t}_x$, $\tilde{t}_y$). We give also the number of correspondences, $Corr$.