Pixel-Based Hyperparameter Selection for Feature-Based Image Registration

F. Brunet[1,2] and A. Bartoli[1] and N. Navab[2] and R. Malgouyres[3]

[1]ISIT, Université d'Auvergne, Clermont-Ferrand, France
[2]CAMPAR, TU München, Germany
[3]LIMOS, UMR 6158, Clermont-Ferrand, France

Abstract

This paper deals with parametric image registration from point correspondences in deformable environments. In this problem, it is essential to determine correct values for hyperparameters such as the number of control points of the warp, a smoothing parameter weighting a term in the cost function, or an M-estimator threshold. This is usually carried out either manually by a trial-and-error procedure or automatically by optimizing a criterion such as the Cross-Validation score. In this paper, we propose a new criterion that makes use of all the available image photometric information. We use the point correspondences as a training set to determine the warp parameters and the photometric information as a test set to tune the hyperparameters. Our approach is fully robust in the sense that it copes with both erroneous point correspondences and outliers in the images caused by, for instance, occlusions or specularities.

Categories and Subject Descriptors (according to ACM CCS): I.4.3 [Computer Graphics]: Image Processing and Computer Vision—Registration
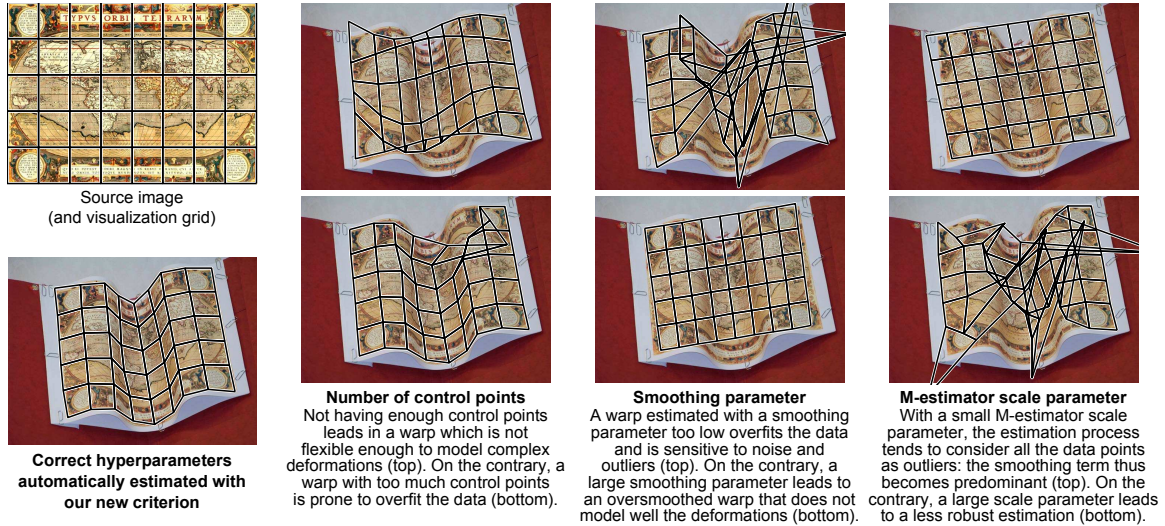
1. Introduction

Parametric image registration is the problem of finding the (natural) parameters of a warp such that it aligns a source image to a target image. In addition to these natural parameters, one also has to determine correct values for the problem *hyperparameters* in order to get a proper registration. The hyperparameters are either additional parameters of the warp itself (*warp hyperparameters*) or parameters included in the cost function to optimize (*cost hyperparameters*). As illustrated in figure 1, the hyperparameters greatly influence the quality of the estimated warp. As reviewed in [Sze06], there are two main approaches to image registration: the feature-based and the pixel-based (or direct) approaches. They both have their own drawbacks and advantages but neither of them directly enables one to automatically tune the hyperparameters. In this paper, we propose a new method to automatically set the hyperparameters by combining the advantages of the feature-based and the pixel-based approaches.

As just said, some hyperparameters are linked to the warp. Let $\mathcal{W} : \mathbb{R}^2 \times \mathbb{R}^l \to \mathbb{R}^2$ be a warp. It is primarily parametrized by a set of l parameters arranged in a vec-

tor $\mathbf{s} \in \mathbb{R}^l$. The homography [HZ04, Sze06] is an example of warp, often parametrized by the 8 independent coefficients of the homography matrix. Another example of warp is the Free-Form Deformation (FFD) [RSH*99] parametrized by $l/2$ two-dimensional control points. Examples of hyperparameters linked to the warps include, but are not limited to, the number of control points of an FFD or the kernel bandwidth of a Radial Basis Function [Boo89].

In the *feature-based approach* [Mod04, Sze06] the source and the target images are 'abstracted' by a finite set of features. In this paper, we consider point features only. The points are then matched to build a set of point correspondences $\{\mathbf{p}_i \leftrightarrow \mathbf{q}_i\}_{i=1}^n$. The basic principle of the feature-based approach is to minimize the error between the warped points of the source image and their corresponding points in the target image. To do so, a robust measure such as an M-estimator can be used (more advanced methods can also be used such as the one described in [PLF08]). Ancillary constraints are often added [Mod04]. For instance, a term promoting smooth warps is generally required to cope with problems such as the lack of accuracy in the location of the

Source image
(and visualization grid)

Number of control points
Not having enough control points
leads in a warp which is not
flexible enough to model complex
deformations (top). On the contrary, a
warp with too much control points
is prone to overfit the data (bottom).

Smoothing parameter
A warp estimated with a smoothing
parameter too low overfits the data
and is sensitive to noise and
outliers (top). On the contrary, a
large smoothing parameter leads to
an oversmoothed warp that does not
model well the deformations (bottom).

M-estimator scale parameter
With a small M-estimator scale
parameter, the estimation process
tends to consider all the data points
as outliers: the smoothing term thus
becomes predominant (top). On the
contrary, a large scale parameter leads
to a less robust estimation (bottom).

**Correct hyperparameters
automatically estimated with
our new criterion**

Figure 1: *Illustration of how some typical hyperparameters influence image registration. The contribution of this paper is a method able to select the proper hyperparameters by combining the advantages of the feature-based and of the pixel-based approaches to image registration. In this example, the data points were automatically detected and matched with SIFT [Low04, VF08]. There was approximately 200 point correspondences (not shown in the figure) uniformly spread across the source image. Among these points, around 10% were gross outliers.*

points or a possible lack of data in some parts of the images. All these elements result in the following minimization problem [Mod04, Bar08]:

$$\min_{\mathbf{s}} \mathcal{E}(\mathbf{s}; \theta), \qquad (1)$$

where θ is a vector containing the hyperparameters and $\mathcal{E}$ is the cost function defined by:

$$\mathcal{E}(\mathbf{s}; \theta) = \sum_{i=1}^{n} \rho\big(\mathcal{W}(\mathbf{p}_i; \mathbf{s}) - \mathbf{q}_i; \gamma\big) + \lambda \mathcal{R}(\mathbf{s}), \qquad (2)$$

with ρ an M-estimator, γ its scale parameter, $\mathcal{R}$ a smoothing term (such as the classical bending energy term discussed in §4) and λ a smoothing term controlling the trade-off between goodness-of-fit and smoothing. In equation (2), γ and λ are two examples of cost hyperparameters. Note that other hyperparameters can appear in the cost function if one decides, for instance, to use more terms. The main advantages of the feature-based approach to image registration are that it copes with large deformations and it is efficient in terms of computational complexity (this is particularly true when using an efficient keypoint detector such as SIFT [Low04] or SURF [BETG08] combined with a good matching algorithm such as the improved nearest neighbour algorithm suggested in [Low04] and implemented in [VF08]). However, the feature-based approach by itself does not enable one to determine correct hyperparameters. As it will be explained in §2, it is *not possible* to determine proper values for the hyperparameters by including them directly in the optimization problem (1), *i.e.* $\min_{\mathbf{s}, \theta} \mathcal{E}(\mathbf{s}; \theta)$.

The other approach to image registration is the *direct approach* [IA99, BM04]. In this case, the warp parameters are estimated by minimizing the pixel-wise dissimilarities between the source image and the warped target image. The main advantage of this approach is that the data used for the parameter estimation is denser than with the feature-based approach. As in the feature-based approach, it is not possible to estimate the hyperparameters with the direct approach.

Since the hyperparameters cannot be trivially estimated, they are often fixed once and for all according to some empirical (and often unreliable) observations. It is also possible to choose them manually with some kind of trial-and-error procedure. This technique is obviously not satisfactory because of its lack of automatism and of foundations. Several approaches have been proposed to tune the hyperparameters in an automatic way. None of them is specific to image registration. They generally minimize a criterion that depends on the hyperparameters and that assesses the 'quality' of the estimated parameters by measuring the ability of the current estimate to generalize to new data. These approaches include, but are not limited to, Akaike Information Criterion [CE02], Mallow's C_P [RS94], Minimum Description Length criterion, and the techniques relying on Cross-Validation scores [Bar08, BPS*03, WW75] (more details are given in §2).

The common characteristic of the previous approaches to automatically select the hyperparameters is that they are problem generic and, as a consequence, they all rely on the point correspondences only. In the particular context

of image registration, another type of data is available: the *photometric information*. We thus propose a new criterion, named the *photometric criterion*, that uses the point correspondences as a training set and the pixel colors as a test set. Another way to put it is to say that our approach combines the two classical approaches to image registration: roughly speaking, the feature-based approach is used to estimate the natural parameters while the pixel-based approach is used for the hyperparameters. Our photometric criterion is more flexible than the previous approaches in the sense that it can handle simultaneously several hyperparameters of different types (for instance, discrete and continuous hyperparameters can be mixed together). Besides, our approach is much more robust to erroneous data (noise and outliers) than previous approaches based on Cross-Validation. Also, it still works when there are only a few point correspondences. Our new criterion is explained in §3 and its ability to properly tune several hyperparameters simultaneously is experimented in §4 with B-spline warps and the Cauchy M-estimator.

Notation. Vectors are denoted using bold fonts (*e.g.* **q**), matrices using sans-serif fonts (*e.g.* M) and scalars in italics (*e.g. x*). The euclidean norm of a vector **v** is written $\|\mathbf{v}\|$. Images are denoted using calligraphic font (*e.g.* $\mathcal{I}$) ; they are considered as functions from $\mathbb{R}^2$ to $\mathbb{R}^c$ where c is the number of channels. Image evaluation at non-integer locations is carried out using bilinear interpolation.

2. Previous Work on Hyperparameter Selection

2.1. Automatic Hyperparameter Estimation

We presented several hyperparameters in the introduction. It is important to understand that inconsistent results would arise if one tries to estimate the hyperparameters by including them in the optimization problem (1). For instance, with such an approach, the best way to minimize the contribution of the regularization term would be to set $\lambda = 0$ which is obviously not the desired value. All the same way, making the M-estimator scale parameter γ tend to 0 would 'artificially' decrease the value of the cost function because it would be equivalent to consider that almost all the point correspondences are outliers (and the cost assigned to outliers tends to zero when $\gamma \to 0$).

The classical approach to build an automatic procedure for selecting the hyperparameters consists in designing a criterion $\mathcal{C}$ that assesses the 'quality' of a given set of hyperparameters [Bar08, Wah90]. The minimizer of this criterion should be the set of hyperparameters to use. The complete problem thus consists in solving the following nested optimization problem:

$$\min_{\mathbf{s}} \mathcal{E}(\mathbf{s}; \arg\min_{\theta} \mathcal{C}(\theta)). \qquad (3)$$

Note that the introduction of the criterion $\mathcal{C}$ makes the prob-

lem (3) completely different from the inconsistent problem $\min_{\mathbf{s}, \theta} \mathcal{E}(\mathbf{s}; \theta)$.

2.2. Cross-Validation

The Cross-Validation (hereinafter abbreviated CV) is a general principle used to tune the hyperparameters in parameter estimation problems [Wah90]. Broadly speaking, a CV procedure consists in minimizing a score function that measures how well a set of estimated parameters will generalize to new data. This is achieved by dividing the whole data set into several subsets. Each one of these subsets is then alternatively used as a training set or as a test set to build the CV score function. The use of CV to select the hyperparameters for spline parameter estimation has been introduced in [WW75]. It has been successfully applied for deformable warp estimation from point correspondences in [Bar08]. We now present two variants of CV: the *Ordinary CV* and the *V-fold CV*.

Ordinary CV (OCV). For a given set of hyperparameters θ, let $\mathbf{s}_\theta^{(k)}$ be the warp parameters estimated from the data with the k-th point correspondence left out. The OCV score, denoted $\mathcal{C}_{OCV}$, is defined by:

$$\mathcal{C}_{OCV}(\theta) = \frac{1}{n} \sum_{k=1}^{n} \left\| \mathbf{q}_k - \mathcal{W}\left(\mathbf{p}_k; \mathbf{s}_\theta^{(k)}\right) \right\|^2. \qquad (4)$$

Tuning the hyperparameters using the OCV consists in minimizing $\mathcal{C}_{OCV}$ with respect to θ. This approach has several drawbacks. First, computing $\mathcal{C}_{OCV}$ is prohibitive: evaluating $\mathcal{C}_{OCV}$ for a single θ with formula (4) requires to estimate each one of the n vectors $\{\mathbf{s}_\theta^{(k)}\}_{k=1}^{n}$. There exists some close approximations of (4) resulting in a significant improvement in terms of computational time. However, these approximations are only usable in a least-squares framework for parameter estimation (see, for instance, [Bar08, FBM08]). Second, the score $\mathcal{C}_{OCV}$ is not robust to false point correspondences. And last, but not least, the OCV score is not reliable when there are not enough point correspondences [Wah90].

V-Fold Cross-Validation (V-fold CV). An alternative to the OCV score is the V-fold CV score. A complete review of the V-fold CV is given in [BPS*03]. It consists in splitting the set of point correspondences into V disjoint sets of nearly equal sizes (with V usually chosen as $V = \min(\sqrt{n}, 10)$). Let $\mathbf{s}_\theta^{[v]}$ be the warp parameters obtained from the data with the v-th group left out and let m_v be the number of point correspondences in the v-th group. The V-fold CV score, denoted $\mathcal{C}_V$, is defined by:

$$\mathcal{C}_V(\theta) = \sum_{v=1}^{V} \frac{m_v}{n} \sum_{k=1}^{m_v} \frac{1}{m_v} \left\| \mathbf{q}_k - \mathcal{W}\left(\mathbf{p}_k; \mathbf{s}_\theta^{[v]}\right) \right\|^2. \qquad (5)$$

The V-fold CV is not robust to erroneous point correspondences. It can be made robust by replacing the average

$\sum_{k=1}^{m_v} \frac{1}{m_v} \left\| \mathbf{q}_k - \mathcal{W}\left(\mathbf{p}_k; \mathbf{s}_\theta^{[v]}\right) \right\|^2$ in equation (5) with some robust measure such as the trimmed mean [BPS*03]. Besides, the V-fold CV score is not more reliable than the OCV score when there are only a few point correspondences.

2.3. Other Approaches

Other approaches such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Mallow's C_P, Minimum Description Length (MDL) have been used to tune hyperparameters (see, for instance, [BPS*03, CE02]). Some robust versions also exist for these criteria ; for instance a robust Mallow's C_P is developed in [RS94]. However, these criteria have usually been developed to choose one model among a finite set of given models and, as such, approaches based on CV are better suited to tune continuous hyperparameters [Bar08].

3. Our Contribution: the Photometric Error Criterion

The common characteristic of the approaches reviewed in §2 is that both the parameters and the hyperparameters are estimated using exactly the same data set, *i.e.* the point correspondences. In this section, we propose a new criterion to tune hyperparameters that makes use of all the available information: not only the point correspondences but also the photometric information.

The principle of our approach consists in combining the two standard approaches to image registration:

- given a set of hyperparameters θ, the feature-based approach is used to determine the warp parameters $\mathbf{s}_\theta$ from the point correspondences ;
- the cost function of the direct approach is used to assess the correctness of the hyperparameters θ: the proper hyperparameters must be the ones minimizing the pixel-wise photometric discrepancy between the target image and the warped source image.

In other words, we propose to use the point correspondences as the training set and the photometric information as the test set. Dividing the data into a training set and a test set is a classical approach of statistical learning [HTF03]. Given a vector of hyperparameters θ and the corresponding warp parameters $\mathbf{s}_\theta$ (estimated from the point correspondences), our criterion, denoted $\mathcal{C}_\star$, is defined as:

$$\mathcal{C}_\star(\theta) = \frac{1}{|\mathfrak{R}|} \sum_{\mathbf{p} \in \mathfrak{R}} \left\| \mathcal{S}(\mathbf{p}) - \mathcal{T}(\mathcal{W}(\mathbf{p}; \mathbf{s}_\theta)) \right\|^2, \quad (6)$$

where $\mathfrak{R}$ is the region of interest and $|\mathfrak{R}|$ its size. $\mathfrak{R}$ can be defined as, for instance, a rectangle obtained by cropping the domain of the source image. $\mathcal{S}$ and $\mathcal{T}$ denote the source and the target images respectively.

Note that the criterion of equation (6) is the cost function typically minimized in direct image registration [IA99,

Sze06]. The difference with direct image registration is that the criterion is considered as a function of the hyperparameters θ, *not* of the warp parameters $\mathbf{s}$.

Robustness. When using photometric information, one should take care of the fact that there can be outliers in the image colors caused, for instance, by occlusions or specularities. The criterion $\mathcal{C}_\star$ can be made robust to these outliers by replacing the squared Euclidean norm in equation (6) with a more robust measure such as the trimmed mean. We thus define the *robust photometric error criterion*, denoted $\mathcal{C}'_\star$, as:

$$\mathcal{C}'_\star(\theta) = \frac{1}{\frac{100-\alpha}{100}|\mathfrak{R}|} \sum_{\mathbf{p} \in \mathfrak{R}_\alpha} \left\| \mathcal{S}(\mathbf{p}) - \mathcal{T}(\mathcal{W}(\mathbf{p}; \mathbf{s}_\theta)) \right\|^2, \quad (7)$$

where $\mathfrak{R}_\alpha$ is the subset of $\mathfrak{R}$ obtained by removing from $\mathfrak{R}$ the α% of the pixels that produce the highest values for $\left\| \mathcal{S}(\mathbf{p}) - \mathcal{T}(\mathcal{W}(\mathbf{p}; \mathbf{s}_\theta)) \right\|^2$.

4. Experimental Results

4.1. Technical Details

In this section, we instantiate our general contribution in order to conduct some experiments.

Warp. The warp we use is the Free-Form Deformation model relying on tensor-product B-Splines, as in [RSH*99]. This warp is parameterized by a set of $l/2$ two-dimensional control points $\mathbf{s}_{ij}$; $i \in \{1, \dots, l_x\}$, $j \in \{1, \dots, l_y\}$ with $l_x l_y = l/2$. They are arranged in a vector $\mathbf{s} \in \mathbb{R}^l$. For a point $\mathbf{p} = (x, y)$, the FFD warp is defined by:

$$\mathcal{W}(\mathbf{p}; \mathbf{s}) = \sum_{i=1}^{l_x} \sum_{j=1}^{l_y} \mathbf{s}_{ij} N_i(x) N_j(y). \quad (8)$$

The values l_x and l_y are two hyperparameters determining the number of control points along the x-axis and the y-axis respectively. The functions N_i are the B-spline basis functions [Die93, RSH*99, dB01] which are polynomials of degree 3. If point $\mathbf{p}$ is fixed and known then the warped point $\mathcal{W}(\mathbf{p}; \mathbf{s})$ is expressed as a linear combination of the control points $\mathbf{s}_{ij}$, and hence can be written in the form $\mathcal{W}(\mathbf{p}; \mathbf{s}) = \mathbf{w}_\mathbf{p}^\mathsf{T} \mathsf{S}$, where $\mathbf{w}_\mathbf{p} \in \mathbb{R}^l$ is a vector depending only on the point $\mathbf{p}$ and $\mathsf{S} \in \mathbb{R}^{l/2 \times 2}$ is the matrix obtained by stacking the control points $\mathbf{s}_{ij}$ (S is a rearrangement of $\mathbf{s}$).

Smoothing Term. In our experiments, the smoothing term $\mathcal{R}$ in equation (1) is replaced by the classical *bending energy*:

$$\mathcal{R}(\mathbf{s}) = \sum_{i=1}^{2} \int_\Omega \left\| \frac{\partial^2 \mathcal{W}^i}{\partial \mathbf{p}^2}(\mathbf{p}; \mathbf{s}) \right\|_\mathcal{F}^2 d\mathbf{p}, \quad (9)$$

where Ω is the domain on which the warp $\mathcal{W}$ is defined, $\mathcal{W}^i$ is the i-th coordinate of the warp, and $\| \cdot \|_\mathcal{F}$ is the Frobenius norm of the Hessian matrix. With FFD, there exists a closed-form expression for the bending energy: $\mathcal{R}(\mathbf{s}) =$

$\mathbf{s}^\mathsf{T}\mathbf{B}\mathbf{s}$, where $\mathbf{B} \in \mathbb{R}^{l \times l}$ is a symmetric, positive, and semi-definite matrix which can be easily computed from the second derivatives of the B-spline basis functions.

M-estimator. In this section we use the Cauchy M-estimator defined by the following ρ function:

$$\rho(x;\gamma) = \log\left(1 + \frac{x^2}{\gamma^2}\right), \qquad (10)$$

where $\gamma \in \mathbb{R}_+^*$ is an hyperparameter that controls the scale of this M-estimator. It can easily be shown that the Cauchy M-estimator is the negative likelihood with errors following a Cauchy/Lorentz distribution. The inaccuracies of the keypoints' locations detected by SURF and SIFT tend to follow such a distribution. Besides, the probability density function (PDF) of the Cauchy distribution has heavy tails that satisfactorily models the outliers, *i.e.* the false point correspondences. We report in figure 2 an illustrative test showing that assuming a Cauchy distribution is consistent with the kind of errors encountered in real cases. In this experiment, we use the source and the target images of figure 1 for which the ground truth warp is known (manually determined). Figure 2 depicts an histogram of the errors between the location of the 1112 keypoints detected with SIFT in the target image and their expected location (computed by applying the ground truth warp to the keypoints in the source image). It shows that considering a Cauchy distribution is a reasonable choice. In particular, the fact that the tails of the PDF of the Cauchy distribution are heavier than the ones of, for instance, the Gaussian PDF makes the cost function of equation (1) robust to outliers.
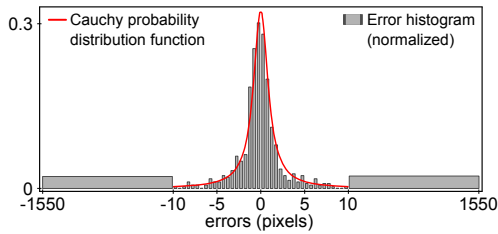

Figure 2: *Graphical comparison between the probability density function of the Cauchy distribution and the (normalized) histogram of the errors between the expected keypoints in the target image and the keypoints automatically detected with SIFT. Mind the scale of the abscissa axis.*

Optimization of the Criteria. All the criteria used in the experiment (including the CV criteria and our new criterion) are minimized using an exhaustive search approach. It consists in evaluating the criteria over a fine grid in order to find the optimum. Although long to compute, this approach has the advantage of being reliable. Besides, we generally optimize over only 2 or 3 hyperparameters, which makes the computational time reasonable.

4.2. Synthetic Data

In this subsection, several experiments are done on synthetic data. Using such data is interesting since it allows us to know precisely the *ground truth* warp that relates the source and the target images.

Synthetic Data Generation. A pair of images is generated from a texture image (randomly chosen in a stock of 15 different images). A rectangular part of the texture image is used as the source image. The target image is build by deforming another part of the texture image with a ground truth warp $\mathcal{W}^\star$, as illustrated in figure 3. The warp $\mathcal{W}^\star$ is a B-spline with 5×5 control points determined randomly and such that the average deformation magnitude is approximately 20 pixels. The sizes of the source and of the target images are 160×160 pixels and 320×240 pixels respectively. A Gaussian noise with standard deviation equal to 5% of the maximal intensity value is added to the pixels of both the source and the target images. A set $\mathcal{P} = \{\mathbf{p}_i \leftrightarrow \mathbf{q}_i\}_{i=1}^n$ of point correspondences is built by randomly picking the points $\mathbf{p}_i$ in the source image and computing their correspondents $\mathbf{q}_i$ in the target image with the warp $\mathcal{W}^\star$. A Cauchy noise with scale parameter $\gamma = 1$ pixel is added to the point correspondences.
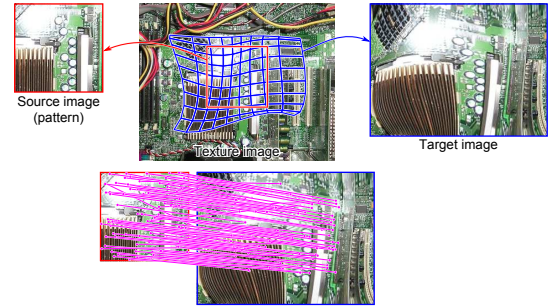
Figure 3: *Synthetic data generation process.*

Oracle. We call *oracle* the warp estimated from the point correspondences $\mathcal{P}$ which is as close as possible to the ground truth warp $\mathcal{W}^\star$. It is designed to be the best possible warp given *i)* the available data and *ii)* the warp model. It is preferable to use the oracle instead of the ground truth warp to evaluate an estimated warp. Indeed, en error between an estimated warp and the ground truth warp does not necessarily comes from a bad estimation process (which is the object of our experiments in this paper): it can comes from the fact that the considered warp model is simply not able to fit the ground truth warp (for example, even if the correct hyperparameters are given, a homography will never fit a highly deformed warp). The oracle is defined as the warp induced by the parameters and the hyperparameters $(\mathbf{s}_o, \theta_o)$ estimated by solving the following problem:

$$(\mathbf{s}_o, \theta_o) = \arg\min_{(\mathbf{s},\theta)} \iint_{\mathbf{p} \in \Omega_{\mathcal{W}^\star}} \|\mathcal{W}^\star(\mathbf{p}) - \mathcal{W}(\mathbf{p};\mathbf{s})\| \, \mathrm{d}\mathbf{p}. \quad (11)$$

Problem (11) is numerically solved using an exhaustive search approach.

4.2.1. Relative Geometric Error (RGE)

The RGE measures the discrepancy between an estimated warp and the oracle. Let $\theta_\bullet$ be the set of hyperparameters that minimizes the criterion $\mathcal{C}_\bullet$ (the symbol $\bullet$ is a place-holder for the criterion name). Let $\mathbf{s}_\bullet$ be the warp parameters estimated from the point correspondences with the hyperparameters $\theta_\bullet$. The RGE is defined as:

$$\iint_{\mathbf{p} \in \Omega_\mathcal{S}} \frac{\|\mathcal{W}(\mathbf{p};\mathbf{s}_o) - \mathcal{W}(\mathbf{p};\mathbf{s}_\bullet)\|}{\|\mathcal{W}(\mathbf{p};\mathbf{s}_o)\|} \, d\mathbf{p}. \tag{12}$$

Figure 4 compares the RGE obtained by tuning the M-estimator scale parameter γ and the smoothing parameter λ with different approaches:

- our photometric criterion (Photo) and its robust versions with thresholds for the trimmed mean of 25% (PhotoR25) and 50% (PhotoR50) ;
- the V-fold CV criterion (VFold) and its robust versions with thresholds for the trimmed mean of 20% (VFoldR20) and 40% (VFoldR40).

The number of control points of the warp is set to 8×8. 100 point correspondences are used to estimate the warp. The results reported in figure 4 are averaged over 100 trials (with different texture images, different point correspondences, and different deformations).

We can observe in figure 4 that the smallest RGE are consistently obtained with our photometric criterion. The difference between robust and non-robust versions of our criterion is not as significant as for the CV criteria. This comes from the fact that in the synthetic data used for this experiment, there are outliers in the point correspondences (thus affecting the non-robust CV scores) while the source and the target images are outlier-free.

4.2.2. Scale Parameter of the Cauchy's M-estimator

Figure 5 shows the values determined with several criteria for the Cauchy's M-estimator scale parameter γ. In addition to the criteria used in the previous experiment, we also show the results obtained with the oracle. The data used in this experiment are the same than the one used in the previous experiment. The point correspondences were generated with errors following a Cauchy distribution with scale parameter equals to 1. As a consequence, the criteria are expected to give the value 1 for the scale parameter of the Cauchy's M-estimator. Figure 5 shows that the proposed photometric criteria results in values for γ which are close to 1. We observe that the three approaches based on the basic V-fold CV also results in correct values. On the contrary, the robust variants of the V-fold CV gives values farther away from 1 than the other approaches. The fact that the value 1 is not exactly retrieved with our criteria is not really significant since this value is not precisely retrieved with the oracle itself.
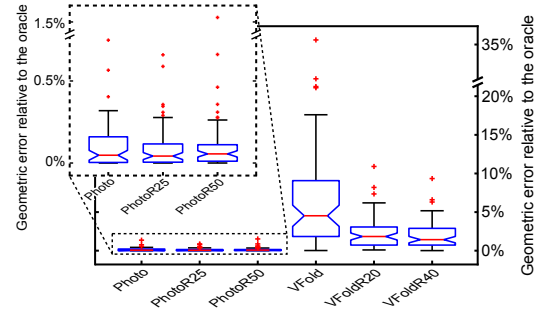

Figure 4: *Relative geometric errors for several criteria used to determine hyperparameters. Globally, the criteria we propose (Photo, PhotoR25, and PhotoR50) give better results than the ones obtained with criteria relying on Cross-Validation (VFold, VFoldR20, and VFoldR40). The red line is the median over the 100 trials. The limits of the blue box are the 25th and the 75th percentiles. The black 'whiskers' cover approximately 99.3% of the experiment outcomes. The red crosses are the outcomes considered as outliers.*
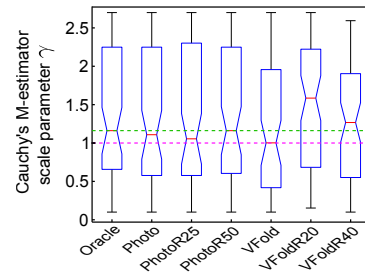
Figure 5: *Scale parameter of the Cauchy's M-estimator retrieved using several criteria. The pink dashed line represents the expected value for this hyperparameter. The green dashed line represents the value retrieved using the oracle. The use of the criteria we proposed (Photo, PhotoR25, and PhotoR50) results in values close to the expected ones.*

4.2.3. Noise in the Point Correspondences

In this experiment, we study the influence of the noise in the point correspondences. We use the same data than in the experiments of §4.2.1 except that there are no outliers in the images. The point correspondences are perturbed using an additive Gaussian noise of standard deviation σ varying between 0 and 12 pixels. Therefore, we only test the non-robust methods: VFold and Photo. These methods are used to automatically tune the regularization parameter. Figure 6 shows the evolution of the RGE in function of the amount of noise in the point correspondences. It shows that our approach Photo is much more robust to the noise than VFold. This comes from the fact that VFold entirely relies on the noisy point correspondences while our approach also includes color information.
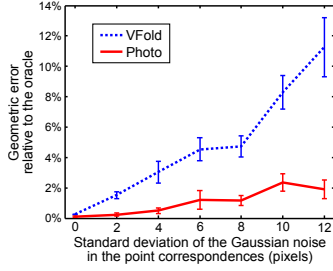
Figure 6: *Evolution of the relative geometric error in function of the (Gaussian) noise in the point correspondences. Our approach,* Photo, *is more robust than the approach relying on the CV (* VFold*).*

4.3. Real Data

The last experiments of this paper are conducted on real data. The source images are digital pictures. The target images are obtained by first printing the source images and second picturing them with a standard camera. Ground truth warps were determined manually by clicking several hundreds of point correspondences in the images. Note that figure 1 shows an example of our approach applied to real data.

4.3.1. The cubist image

Figure 7 shows the registration results obtained by automatically determining the hyperparameters with several criteria. In this experiment, three hyperparameters were considered: the smoothing parameter λ, the M-estimator threshold γ, and the number of control points of the B-spline warp l_x (the number of control points along the x-axis and the y-axis were set to be equal). 314 point correspondences were automatically determined using the SIFT detector and the descriptor matcher implemented in [VF08]. Approximately 8% of the point correspondences were false matches. We can observe in figure 7 that our photometric criterion is the one giving the best results. The standard V-Fold CV criterion is the one leading to the worst results due to the presence of erroneous point correspondences. The robust V-Fold CV criterion performs better than the non-robust one but is not as good as ours, particularly for the bottom right corner of the image: this is due to a lack of point correspondences in this part of the image.

We report in table 1 the RGE as defined in §4.2.1 for the warps estimated in the 'cubist image' experiment.

Criterion	RGE
V-Fold CV	1.852%
V-Fold CV (robust)	0.675%
Our criterion	0.190%
Our criterion (robust variant)	0.197%

Table 1: *RGE for the experiment of figure 7.*
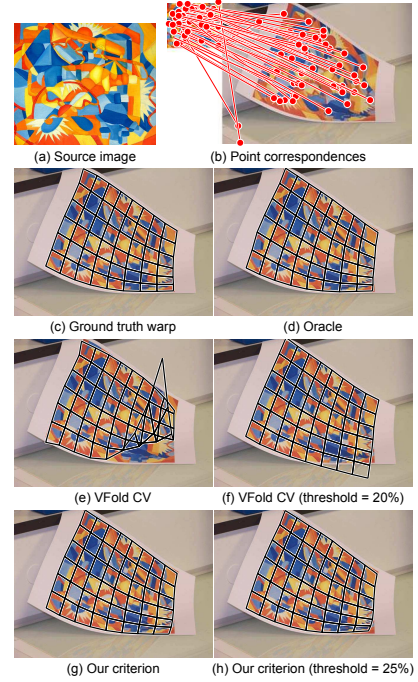
Figure 7: *Image registered with 3 hyperparameters (γ, λ, and l) automatically determined with several criteria. The point correspondences were obtained with SIFT. The thresholds indicated in (f) and (h) are the thresholds of the trimmed means (see §2.2 and §3). In this case, the two variants of our criterion are the ones that lead to the best results.*

4.3.2. 'Waterfall' of Maurits Escher

Figure 8 shows an experiment similar to the one conducted with the 'cubist image'. Nonetheless, there are some important differences. This time, the keypoints were extracted using the SURF detector of [BETG08] and approximately 12% of the 621 point correspondences were erroneous. An artificial occlusion was added to the target image; we used an artificial occlusion in order to still be able to determine the ground truth warp (which is done before the insertion of the occlusion). Besides, the M-estimator scale parameter and the smoothing parameter were the only hyperparameters under study (the number of control points of the warp was set to the one of the ground truth warp). As in the 'cubist image' case, the hyperparameters chosen with our photometric criterion are better than the ones estimated with the criterion relying on the V-Fold Cross-Validation. In both cases, the robust versions of the criteria perform better than the non-robust ones. Note that the occlusion added to the target image influences the non-robust V-fold CV criterion since it introduces supplementary false point correspondences.

5. Conclusion

We proposed a new criterion to automatically tune the hyperparameters in image registration problems. We showed
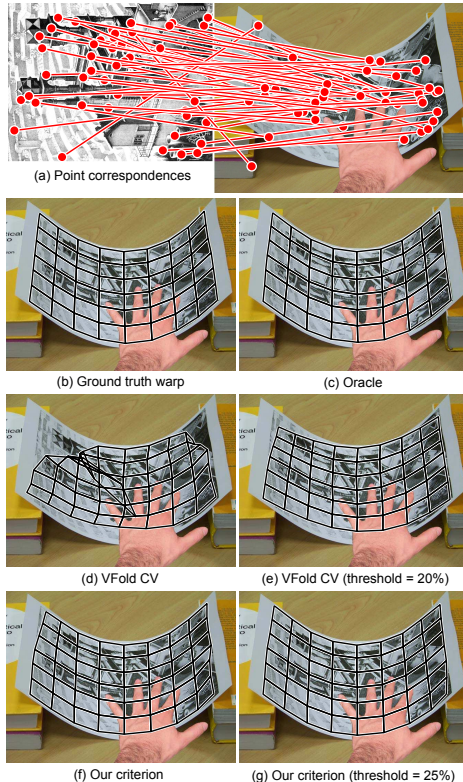
Figure 8: *Image registered with 2 hyperparameters (γ, λ) automatically determined with several criteria. The point correspondences were obtained with SURF. The thresholds indicated in (f) and (h) are the thresholds of the trimmed means (see §2.2 and §3). Globally, the robust variants of the VFold CV criterion and of our criterion lead to acceptable results. The non-robust VFold CV criterion is greatly influenced by the presence of outliers in the point correspondences. The non-robust variant of our criterion is slightly more influenced by the occlusion than the robust variant.*

that our photometric criterion performs generally better than other approaches with similar goals such as the Cross-Validation criteria. This was made possible by designing a criterion specifically adapted to the image registration problem that combines the advantages of both the feature-based and the pixel-based approaches to image registration. Our criterion was successfully experimented in a particular but challenging setup: deformable B-spline warps, selection of an M-estimator threshold, presence of outliers and occlusions, *etc.* However, the proposed criterion is not limited to this setup: it is generic enough to be applied in other image registration problems with different constraints, different warps, and, thus, different hyperparameters.

References

[Bar08] BARTOLI A.: Maximizing the predictivity of smooth deformable image warps through cross-validation. *Journal of Mathematical Imaging and Vision 31*, 2-3 (2008), 133–145. 2, 3, 4

[BETG08] BAY H., ESS A., TUYTELAARS T., GOOL L. V.: SURF: Speeded up robust features. *Computer Vision and Image Understanding 110* (2008), 346–359. 2, 7

[BM04] BAKER S., MATTHEWS I.: Lucas-Kanade 20 years on: A unifying framework. *International Journal of Computer Vision 56* (2004), 221–255. 2

[Boo89] BOOKSTEIN F.: Principal warps: Thin-Plate Splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11* (1989), 567–585. 1

[BPS*03] BRABANTER J. D., PELCKMANS K., SUYKENS J., VANDEWALLE J., MOOR B. D.: *Robust Cross-Validation Score Functions with Application to Weighted Least Squares Support Vector Machine Function Estimation*. Tech. rep., Katholieke Universiteit Leuven, 2003. 2, 3, 4

[CE02] CETIN M., ERAR A.: Variable selection with Akaike information criteria: a comparative study. *Hacettepe Journal of Mathematics and Statistics 31* (2002), 89–97. 2, 4

[dB01] DE BOOR C.: *A Practical Guide to Splines – Revised Edition*. Springer, 2001. 4

[Die93] DIERCKX P.: *Curve and Surface Fitting with Splines*. Oxford University Press, 1993. 4

[FBM08] FARENZENA M., BARTOLI A., MEZOUAR Y.: Efficient camera smoothing in sequential structure-from-motion using approximate cross-validation. In *European Conference on Computer Vision* (2008). 3

[HTF03] HASTIE T., TIBSHIRANI R., FRIEDMAN J. H.: *The Elements of Statistical Learning*. Springer, 2003. 4

[HZ04] HARTLEY R., ZISSERMAN A.: *Multiple View Geometry in Computer Vision*, second ed. Cambridge University Press, 2004. 1

[IA99] IRANI M., ANANDAN P.: About direct methods. In *Workshop on Vision Algorithms* (1999). 2, 4

[Low04] LOWE D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60* (2004), 91–110. 2

[Mod04] MODERSITZKI J.: *Numerical Methods for Image Registration*. Oxford Science, 2004. 1, 2

[PLF08] PILET J., LEPETIT V., FUA P.: Fast non-rigid surface detection, registration and realistic augmentation. *International Journal of Computer Vision 76*, 2 (2008). 1

[RS94] RONCHETTI E., STAUDTE R.: A robust version of Mallow's C_P. *Journal of the American Statistical Association 89*, 426 (1994), 550–559. 2, 4

[RSH*99] RUECKERT D., SONODA L., HAYES C., HILL D., LEACH M., HAWKES D.: Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging 18* (1999), 712–721. 1, 4

[Sze06] SZELISKI R.: Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Vision 2* (2006), 1–104. 1, 4

[VF08] VEDALDI A., FULKERSON B.: VLFeat: An open and portable library of computer vision algorithms. http://www.vlfeat.org/, 2008. 2, 7

[Wah90] WAHBA G.: *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, 1990. 3

[WW75] WAHBA G., WOLD S.: A completely automatic French curve: fitting spline functions by cross-validation. *Commun. Stat. 4* (1975), 1–17. 2, 3