p-adic methods in stereo vision

Patrick Erik Bradley

Patrick Erik Bradley Institut für Photogrammetrie und Fernerkundung, Karlsruhe Institut für Technologie (KIT), Englerstr. 7, 76128 Karlsruhe, Germany bradley@kit.edu

Abstract. The so-called *essential matrix* relates corresponding points of two images from the same scene in 3D, and allows to solve the relative pose problem for the two cameras up to a global scaling factor, if the camera calibrations are known. We will discuss how *Hensel's lemma* from number theory can be used to find geometric approximations to solutions of the equations describing the essential matrix. Together with recent p-adic classification methods, this leads to RanSaC$_p$, a p-adic version of the classical RANSAC in stereo vision. This approach is motivated by the observation that using p-adic numbers often leads to more efficient algorithms than their real or complex counterparts.

1 Introduction

According to [9], ultrametricity is pervasive in observational data, and this offers computational advantages and a well understood basis for developping data processing tools originating in p-adic arithmetic. Consequently, p-adic data encoding becomes necessary. In [1] it has been shown that the choice of the prime number p is arbitrary. Hence $p = 2$ can be taken, which is usually the computationally most advantageous prime number. In particular, the p-adic Newton iteration method, known in number theory as Hensel's lemma, is most efficient for $p = 2$. We will use this method in order to give way to computationally efficient methods for solving the relative pose problem from five corresponding points in stereo vision.

A well known hierarchical image encoding procedure is the (regular) *quadtree*. We will show that it has natural 2-adic encodings which allow to view gray-scale images as real-valued functions on p-adic spaces. This should be understood as an invitation to develop image processing methods originating in p-adic functional analysis. In any case, image coordinates are p-adic numbers in this situation. Although computationally efficient, the quadtree suffers somewhat from its rigidity when it comes to handling measurement errors. We expect that taking families of 2-adic encodings corresponding to small euclidean perturbations will lead to a dynamic treatment of single images which can overcome this drawback without losing too much computational efficiency.

An Introduction to p-adic numbers is e.g. [5].

2 p-adic numbers

Kurt Hensel's important contribution to number theory was to view numbers as analytic functions on some imagined "Riemann surface". In this imaginary situation, the "places" are given by the prime numbers p which play the role of a local coordinate[1], and then the number n has "locally" a unique power series expansion

$$n = \sum_{\nu=0}^{\infty} n_\nu p^\nu,$$

which in the case of natural numbers n is in fact a finite expansion with coefficients $n_\nu \in \{0, \ldots, p-1\}$. The p-adic metric is given by the length of the common initial part:

$$|n-m|_p = p^{-\nu}, \tag{1}$$

if $m = n_0 + \cdots + n_{\nu-1} p^{\nu-1} + m_\nu p^\nu + \ldots$ and $m_\nu \neq n_\nu$. This is an ultrametric, i.e. the strict triangle inequality

$$|x+y|_p \leq \max\left\{|x|_p, |y|_p\right\}$$

holds true. Allowing infinite expansions (1) means completion with respect to the p-adic metric, and the completed space $\mathbb{Z}_p$ of *p-adic integers* contains the usual integers $\mathbb{Z}$ as a dense subset. Examples of negative numbers are

$$\sum_{\nu=0}^{\infty} p^\nu = \frac{1}{1-p}, \qquad \sum_{\nu=0}^{\infty} (p-1) p^\nu = -1$$

The primality of p guarantees that there are no zero-divisors in $\mathbb{Z}_p$, and the field of fractions $\mathbb{Q}_p$ can be formed which densely contains the rational numbers $\mathbb{Q}$. Just like in the function-theoretic case, the p-adic numbers thus correspond to the meromorphic functions:

$$\mathbb{Q}_p = \left\{ \sum_{\nu=-N}^{\infty} x_\nu p^\nu \mid x_\nu \in \{0, \ldots, p-1\} \right\}$$

and have a "Laurent series" expansion. Observe further that $\mathbb{Z}_p$ is the p-adic unit disk:

$$\mathbb{Z}_p = \left\{ x \in \mathbb{Q}_p \mid |x|_p \leq 1 \right\},$$

and we have in $\mathbb{Q}_p$ an ultrametric space on which calculus can be performed.

p-adic approximation is given by finite expansions: $x = x_0 + \ldots x_{n-1} p^{n-1} +$ higher order terms. That cut-off can be written by a congruence

$$x \equiv x_0 + \cdots + x_{n-1} p^{n-1} \mod p^n, \tag{2}$$

from wich it follows that the p-adic expansion of x is given by an infinite sequence of congruences (2) with $n = 1, 2, 3, \ldots$. And indeed,

$$\left| x - \sum_{\nu=0}^{n-1} x_\nu p^\nu \right|_p \leq p^{-n},$$

[1] In fact, this dream became true thanks to Grothendieck's concept of *scheme*: The "Riemann surface" is the affine scheme $\operatorname{Spec}\mathbb{Z}$, the space whose points are the prime ideals $p\mathbb{Z}$ for $p = 0$ or a prime number. Cf. e.g. [7]

we have convergence of these finite expansions to x for $n \to \infty$.

At last we remark that $\mathbb{Q}_p$ is endowed with a Haar measure dx such that

$$\int\limits_{\mathbb{Z}_p} \mathrm{d}x = 1,$$

i.e. the unit disk has volume 1.

3 p-adic encoding of images

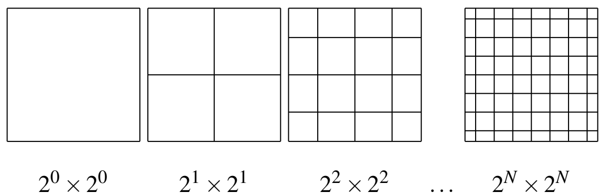

Fig. 1. Hierarchical subdivision of an image

A 2-adic encoding of square $2^N \times 2^N$-images can be obtained by a hierarchical subdivision as in Fig. 1. Essentially, there are two approaches for the encoding. In the *bottom-up encoding*, the squares at highest resolution are assigned to level N, with decreasing level at higher hierarchy, level 0 representing the full image cluster. The encoding scheme for the x-coordinate is to traverse a path from bottom to top, and collect a coefficient $a_\nu = 0$ for each right turn, and $a_\nu = 1$ for left turns. This yields the expansion

$$x = \sum_{\nu=0}^{N} a_\nu 2^{-\nu}.$$

Fig. 2 (left) exemplifies this with

$$x_1 = 0, \qquad x_2 = 2^{-1}, \qquad x_3 = 2^{-2}, \qquad x_4 = 2^{-2} + 2^{-1}.$$

The intensities (gray values) on the image grid can be viewed as *locally constant* functions $f\colon \mathbb{Q}_p \to \mathbb{R}$, as vertices in the dendrogram at level ν can be viewed as p-adic disks of radius p^ν. With the bottom-up encoding the functions are constant on all translates of the unit disk $\mathbb{Z}_p$, and methods from p-adic functional analysis are ready for application. The functions which are constant on the sets $x + \mathbb{Z}_p$ are in one-one correspondence with functions on the co-set space $\mathbb{Q}_p/\mathbb{Z}_p$.

Example 3.1 (**p-adic diffusion**) *p-adic diffusion can be described as a symmetric jump process on $\mathbb{Q}_p/\mathbb{Z}_p$, with the transition probability P_{xy} depending only on their p-adic distance:*

$$P_{xy} = \rho(|x-y|_p).$$

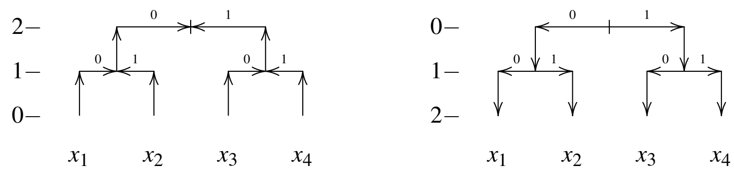

Fig. 2. Left: bottom-up encoding, right: top-down encoding of a dendrogram

The equation describing the evolution of probabilities is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} f(x,t) = \int_{\mathbb{Q}_p} (f(y,t) - f(x,t))\, \rho(|x-y|_p)\, \mathrm{d}y.$$

By taking as integral kernel the function

$$\rho(|x-y|_p) = \frac{|x-y|_p^{-1-\alpha}}{\Gamma_p(-\alpha)}$$

with $\alpha > 0$*, one obtains p-adic Brownian motion. Here, the role of the Laplace operator is played by the* Vladimirov operator

$$D^\alpha f(x,t) = \frac{1}{\Gamma_p(-\alpha)} \int_{\mathbb{Q}_p} \frac{f(x) - f(y)}{|x-y|_p^{1+\alpha}}\, \mathrm{d}y,$$

and $\Gamma_p(\alpha) = \frac{1-p^{\alpha-1}}{1-p^{-\alpha}}$ *is the p-adic gamma function. The diffusion equation*

$$\frac{\mathrm{d}}{\mathrm{d}t} f(x,t) = -D^\alpha f(x,t)$$

can be viewed as the p-adic analogue of a scale-space equation, where t plays the role of the scaling parameter. However, a p-adic scale-space theory, from which feature detectors and descriptors can be derived, has yet to be developped. More on p-adic pseudo-differential equations, functional analyis and mathematical physics can be found in [12].

The *top-down encoding* reverses the order of bottom-up, and expansion is in positive powers of 2. This yields 2-adic integers for image coordinates, which turns out useful in the following section. In Fig. 2 (right) one obtains

$$x_1 = 0, \qquad x_2 = 2^1, \qquad x_3 = 2^2, \qquad x_4 = 2^1 + 2^2.$$

4 Epipolar geometry

Epipolar geometry is the geometry of two pinhole cameras [6]. These are described as projective maps $P_L, P_R \colon \mathbb{P}^3 \to \mathbb{P}^2$ between projective spaces. Imagine the two cameras viewing the same point x in a 3-dimensional scene. The plane spanned by x and the two camera centers O_L, O_R is called the *epipolar plane*. This plane cuts out in each of the cameras' image planes a line, called the *epipolar line*. To the projection point $x_L = P_L x$ of x onto the left image corresponds the line through $x_R = P_R x$ and e_R, where e_R is the intersection of the line $O_L O_R$ with the right image plane. This line is given by

$$(e_R \times x_R)^T x_R = 0.$$

By writing the vector product as multiplication with an anti-symmetric matrix S_R and taking the pseudo-inverse of P_L back to x. More precisely, $e_R \times x_R = S_R x_R$, and $x = P_L^+ x_L$, we obtain

$$0 = (e_R \times x_R)^T x_R = x_R^T (e_R \times x_R) = x_R^T \underbrace{S_R P_R P_L^+}_{=:F} x_L, \tag{3}$$

where F is the *fundamental matrix*. This matrix encodes the relative motion between the two cameras, together with their intrinsic parameters given by the calibration matrices K_L, K_R: $F = K_L^{-1} E K_R$. The matrix E is the *essential matrix* and is the fundamental matrix of two *normalised* cameras, which corresponds to the case where both calibrations are known. The essential matrix decomposes into a rotation R and a translation t: $E = R[t]_\times$, where $[t]_\times$ is the matrix for cross product with t. This decomposition can be effected, and is unique upto a sign ambiguity. The problem of stereo vision is thus reduced to the estimation of E from two given images of the same scene.

The essential matrix is a projective 3×3-matrix. Hence, 8 independent equations of the form (3) suffice for determining E. The first algorithm starts with

$$x_1^T E x_1' = 0, \quad \dots \quad x_8^T E x_8' = 0,$$

where (x_i, x_i') are 8 generic pairs of corresponding points in the two images.

Although the 8-point algorithm became successful, its resulting matrix E is not alwys an essential matrix. This is because further constraints have to be met. For example, from (3) it can be easily seen that an essential matrix must fulfill

$$\det E = 0,$$

which leads to the 7-point algorithm.

The minimal number of point correspondences needed for determining the essential matrix is five. This leads in the generic case to a four-dimensional solution space. Writing the general solution as

$$E = u_1 E_1 + u_2 E_2 + u_3 E_3 + u_4 E_4,$$

one obtains homogeneous linear polynomials in four variables $u_1, \dots, u_4$. The constraints for E are given by

$$E E^T E - \frac{1}{2} \operatorname{Trace}(E E^T) E = 0, \tag{4}$$

equations which describe a space M_E containing all essential matrices [4]. Hence, one is left with solving a system of homogeneous cubic equations in four variables. Demazure showed that there are up to 10 complex solutions to (4). Nistér's first five-point algorithm reduces the system (4) to a univariate equation of degree 10 which then has to be solved numerically [10].

The importance of having the minimal number of point correspondences in order to find finitely many exact solutions to the problem lies in the fact that outliers generally lead to bad results. Higher stability is obtained by a *Random Sample Consensus* (RanSaC). Here, a random sample of 5 pairs of corresponding points in general position is chosen. The up to 10 real candidate essential matrices are collected, and another sample is taken, etc. After sufficiently many samples, the candidate matrix whose ε-neighbourhood contains the highest number of solutions from all samples is declared as the "essential matrix" for the stereo vision problem. In essence, a classification of the candidates is performed, and the biggest cluster contains the winning candidate.

5 Hensel's lemma and RanSaC$_p$

The projective cameras can also be p-adic projective maps $\mathbb{P}^3(\mathbb{Q}_p) \to \mathbb{P}^2(\mathbb{Q}_p)$. Thus we are lead to estimate p-adic essential matrices. The notions of translation and rotation make sense p-adically. The latter is defined algebraically: namely a 3×3-matrix R satisfying

$$RR^T = 1, \quad \det R = 1.$$

The top-down encoding from Section 3 ensures coordinates from $\mathbb{Z}_p$. In particular, the coeffients of (4) are p-adic integers. In this case, Hensel's lifting lemma can be used for p-adically approximating the solutions of (4). The point is that a sequence of linear congruences modulo p^k is iteratively solved, which makes the procedure to a p-adic version of Newton's method. For convenience, we formulate the lemma in the case of n equations in m variables in the most familiar version:

Lemma 5.1 (Hensel) *Let $\mathbf{f}(\mathbf{X}) = (f_1(X_1,\dots,X_m),\dots,f_n(X_1,\dots,X_m))$ be an n-tuple of polynomials in m variables with coefficients from $\mathbb{Z}_p$. Let $\mathbf{a}\in\mathbb{Z}_p^m$ such that*

$$\mathbf{f}(\mathbf{a}) \equiv 0 \mod p, \quad \textit{and} \quad \mathrm{rk}\left(\frac{\partial \mathbf{f}}{\partial \mathbf{X}}(\mathbf{a}) \mod p\right) \quad \textit{is maximal}.$$

Then there is a unique solution $\mathbf{a}'$ of $\mathbf{f}$ near $\mathbf{a}$:

$$\mathbf{f}(\mathbf{a}') = 0, \qquad \mathbf{a}' \equiv \mathbf{a} \mod p.$$

RanSaC$_p$, the p-adic version of Random Sample Consensus, collects all candidate essential matrices from all random samples of five corresponding pairs of image points in a set $\mathscr{E}$, and performs a p-adic classification. The central elements of the largest cluster determine the choice of solution for the problem.

The classification method proposed in [3] is to find a clustering of $\mathscr{E}$ by minimising for fixed k the quantity

$$\varepsilon_p = \varepsilon_p(\mathscr{E},\mathscr{C},\mathbf{a}) = \sum_{C\in\mathscr{C}}\sum_{a\in C}\|a - a_C\|_p,$$

where $\mathscr{C} = \{C_1,\dots,C_\ell\}$, $\ell \le k$ is a clustering of $\mathscr{E}$, $\mathbf{a} = (a_C)_{C\in\mathscr{C}}$ with $a_C\in C$, and

$$\|(x_1,\dots,x_n)\|_p = \max\left\{|x_1|_p,\dots,|x_n|_p\right\}$$

is the maximum norm on $\mathbb{Q}_p^n$ (in the present case $n = 9$). The algorithm is a p-adic adaptation of the classical split-LBG [8], a hierarchical version of k-means by splitting cluster centers in two and regrouping the cluster around the new centers. The p-adic adaptation is called LBG_p and first splits clusters by replacing vertices in the dendrogram by their children, and afterwards finding centers a_C in cluster C which further minimse ε_p (cf. [2] for details).

In the event that there is not a unique biggest cluster, the cluster with highest density can be chosen. This is measured by

$$\delta(C) = \begin{cases} \frac{|C|-1}{\mu(C)}, & |C| > 1 \\ 0, & \text{otherwise} \end{cases}$$

where

$$\mu(C) = \int\limits_{\mathbb{Q}_p^n} 1_{B_C}\, dx$$

with 1_{B_C} the indicator function of the smallest ball in $\mathbb{Q}_p^n$ containing C.

As a further tie-breaking rule can be taken the *cluster precision*, given by

$$\pi(C) = \frac{1}{\mu(C_c)},$$

where C_c is the intersection of C with the smallest ball in $\mathbb{Q}_p$ containing the centers of C.

6 Discussion and outlook

An implementation of RanSaC_p, as outlined in [3], is ongoing work with V. Anashin and his students. Preliminary results are that Hensel lifting is fast whenever possible. At the moment, a number of samples do not fulfill the lifting condition of Lemma 5.1. But, in general, unique lifting is possible also under weaker conditions, and it remains to classify samples according to their liftability behaviour. This leads to the research task of studying the p-adic geometry of Demazure's variety M_E of essential matrices. We expect to be able to identify in M_E the locus of Hensel-liftability, as well as other regions for which some ad-hoc lifting methods can be devised. Ideally, one would be able to read off the 5-point configuration the corresponding region in M_E in order to quickly decide whether to discard the sample or to lift with Hensel or in an ad-hoc manner. This should enable the p-adic 5-point algorithm to outperform its classical counterpart in terms of efficiency, because p-adic numerics are very fast for small primes p (here the choice $p = 2$ becomes an asset).

Another issue is that of registration errors. Namely, erroneous point-correspondences in the two images lead to erroneous essential matrices. The p-adic approach has the drawback that small euclidean inaccuracies can lead to large p-adic errors. The research problem is how to overcome this drawback. A promising idea seems for us to study the variation of the 10 points in M_E given by the translation group $x \to x + \varepsilon$. Namely

$$f(x) \to f(x+\varepsilon) \tag{5}$$

amounts to a shift in the division points for the quadtree-like subdivision underlying the encoding. The error ε is controlled by the (inverse) *Monna map* to the real numbers:

$$\varepsilon = \sum \varepsilon_\nu p^\nu \mapsto M(\varepsilon) = \sum \varepsilon_\nu p^{-\nu-1},$$

where $M(\varepsilon)$ is to be taken small. Here, it is quite tempting to view (5) as part of the Weyl representation of p-adic quantum mechanics [11] which in particular for processing complex-valued images can lead to exciting new p-adic methods.

7 Conclusion

Viewing the hierarchical world as ultrametric leads to the consideration of p-adic methods for detecting and processing hierarchies. For this, p-adic data encoding becomes indispensible. This applied to images yields encodings of special quadtrees, known in image processing. The bottom-up method introduced here opens the way for methods from p-adic mathematical physics, whereas the top-down method renders p-adic integers as image coordinates. The latter allowed the use of Hensel's lifting lemma to the equations arising in the problem of finding the essential matrix from five point-correspondences in stereo vision.

p-adic classification algorithms are known to be more efficient than their classical counterparts. Hence, it is natural to use a recently developed p-adic method as part of RanSaC$_p$, a p-adic form of the Random Sample Consensus applied to the five-point relative pose problem in order to find the "best" p-adic approximation to the essential matrix as the one lying centrally in the biggest cluster.

Acknowledgement. The author thanks Sven Wursthorn and Boris Jutzi for introducing him to the topic of computer vision, and for valuable discussions.

References

1. Bradley, PE (2008) Degenerating families of dendrograms. J Classif 25:27–42
2. Bradley, PE (2009) On p-adic classification. p-Adic Numbers Ultrametric Analysis and Applications 1:271–285
3. Bradley, PE (2010) A p-Adic RANSAC algorithm for stereo vision using Hensel lifting. p-Adic Numbers Ultrametric Analysis and Applications 2:55–67
4. Demazure M (1988) Sur deux probleémes de reconstruction. INRIA Rapports de Recherche 882
5. Gouvêa FQ (1993) p-adic numbers. An introduction. Springer, Berlin
6. Hartley R, Zisserman A (2008) Multiple view geometry in computer vision. Cambridge University Press, Cambridge
7. Hartshorne R (1993) Algebraic geometry. Springer, New York
8. Linde Y, Buzo A, Gray RM (1980) An Algorithm for Vector Quantizer Design. IEEE T Commun 28:84–94
9. Murtagh, F (2004) On ultrametricity, data coding, and computation. J Classif 21:167–184
10. Nistér D (2004) An efficient solution to the five-point relative pose problem. IEEE T Pattern Anal 26:167–184
11. Vladimirov VS, Volovich IV (1989) p-Adic quantum mechanics. Commun Math Phys 123:659–676
12. Vladimirov VS, Volovich IV, Zelenov YeI (1994) p-Adic analysis and mathematical physics. World Scientific, Singapore