Pattern Recognition Letters 34 (2013) 1381–1385

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

A comment on "A novel approach for the registration of weak affine images"

Jan Flusser *, Barbara Zitová

Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Praha 8, Czech Republic

ARTICLE INFO

Article history:
Received 22 December 2012
Available online 17 May 2013
Communicated by G. Borgefors

Keywords:
Image registration
Affine transform
Affine subgroups
Weak affine transform
Area-preserving affine transform

ABSTRACT

In the recent paper (Li and Zhang, 2012) a mapping model of weak affine transform (WAT) was proposed for registering images of flat scenes. We show that such a model may cause inaccuracies when registering images with non-uniform scaling differences, even if they seemingly behave according to this model. Mathematical reason of this is that WAT is neither invertible nor closed. We explain why mapping models forming a group should be preferably used for registration and show some examples of models belonging to affine subgroups.

1. Introduction

Image registration is a process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors in such a way that the corresponding objects should have identical spatial coordinates. Image registration is one of the most important and most frequently discussed topics in image processing measured both by the number of practical applications as well as of the research publications (see Zitova and Flusser, 2003 for a survey). It is a crucial preprocessing step in all image analysis tasks in which the final information is gained from the combination of various data sources (image fusion, change detection, multichannel image restoration, superresolution, etc.).

Traditionally, image registration is a four-stage process: selecting the control point (CP) candidates, matching the candidates, choosing a proper transformation model and calculating its parameters, and, finally, resampling and transformation of the sensed image. Each stage is responsible for specific errors and contributes to overall inaccuracy of the result. This is why the authors working in this domain concentrate on all of them, but namely on the CP matching and the choice of the transform model. These two stages are potential sources of the most serious errors.

In case of 2D non-elastic registration, similarity model (i.e., translation, rotation and uniform scaling) and affine model are the most common ones. Recently, Li and Zhang (2012) proposed a model that they called *weak affine transform* (WAT) and which lies "in between" the similarity and affine models. Li and Zhang recommend to use this model when there is no skewing between the reference and sensed images but the similarity model is insufficient because it cannot handle a non-uniform scaling.

In this paper we demonstrate that the concept of the WAT transform is misleading because it does not fulfil basic requirements of reasonable transformation models. We also show that if there is a necessity of having a model between similarity and affinity, then another concept should be employed.

2. Weak affine transform

Affine transform in 2D is defined as

$$\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{t}, \tag{1}$$

where

$$\mathbf{A} = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \tag{2}$$

is a regular matrix called *affine matrix* or *transform matrix*, $\mathbf{t} = (a_5, a_6)^T$ is a shift vector, and $\mathbf{x} = (x, y)^T$ and $\mathbf{x}' = (x', y')^T$ are the spatial coordinates. Such a transform has six independent parameters and when used as a mapping model, at least three non-collinear CP's are required. As pointed out by Suk and Flusser (2005), any affine transform can be decomposed into shift, rotation, non-uniform scaling, another rotation and possibly mirroring. This corresponds to a singular-value decomposition (SVD) of the matrix **A**:

$$\mathbf{A} = \mathbf{R_1} \cdot \mathbf{S} \cdot \mathbf{R_2}, \tag{3}$$

* Corresponding author. Tel.: +420 26605 2422; fax: +420 2 6641 4903.
E-mail addresses: flusser@utia.cas.cz (J. Flusser), zitova@utia.cas.cz (B. Zitová).

0167-8655/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.patrec.2013.04.024

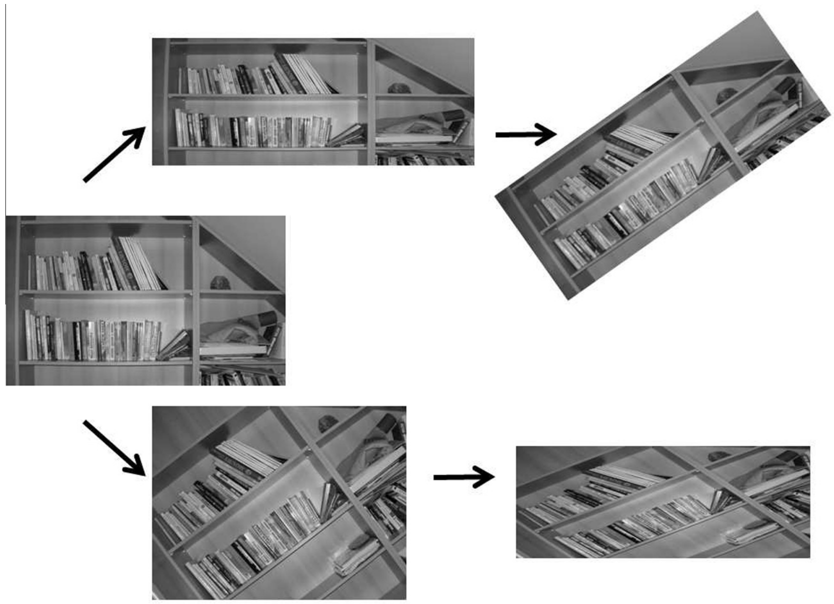

Fig. 1. Weak affine transform is a composition of a non-uniform scaling followed by a rotation (top). A reverse ordering of these two operations leads to a different result (bottom). The bottom row shows what happens in real imaging systems with a non-uniform spatial resolution.

where

$$\mathbf{R_k} = \begin{pmatrix} \cos\alpha_k & -\sin\alpha_k \\ \sin\alpha_k & \cos\alpha_k \end{pmatrix} \quad (4)$$

are two rotation matrices and

$$\mathbf{S} = \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \quad (5)$$

is a matrix of non-uniform scaling.

Li and Zhang (2012) defined *weak affine transform* such that $\mathbf{R_2}$ is just an identity. Thus, the weak affine matrix has the form

$$\mathbf{W} = \mathbf{R_1} \cdot \mathbf{S}, \quad (6)$$

which means the image is first scaled non-uniformly and then rotated (see Eq. 3 in Li and Zhang (2012)). The WAT has five independent parameters which requires the knowledge of at least three non-collinear CP's. The transform parameters are calculated as a least-square solution of an over-determined system.

This WAT model is not reasonably justified from practical point of view. In Li and Zhang (2012) the authors correctly pointed out that the non-uniform scaling is mostly introduced by a different spatial resolution of the sensor in horizontal and vertical directions. However, when the sensor is rotated with respect to the reference coordinates, the operations act in a reverse order: the image is *first* rotated and *then* scaled. These operations are not commutative, the order does matter and the results are different (see Fig. 1 for an example and the next Section for a more detailed discussion). Nevertheless, even if the reverse order was used in the WAT definition, it would not remove the most serious problems with this model that were ignored in Li and Zhang (2012) and that we describe in the next Section.

3. Affine subgroups

The most serious problem with the WAT transform arises from the fact that it does not form a group (subgroup of an affine group). Let us recall that a transform group (subgroup) must be closed under composing (which must be associative), must contain a unit element, and along with any element it must contain also its inverse. Affine transformation fulfil these axioms and form the affine group. Meaningful transform models used in image registration should always form a group.[1] In the opposite case, namely the absence of the closure property and the non-invertibility of the model cause both conceptual as well as practical problems.

The WAT is not closed under composition, the composition of two WAT transforms goes beyond the WAT model. To see that, let us consider two WAT matrices $\mathbf{W} = \mathbf{R_1 S}$ and $\mathbf{W_2} = \mathbf{R_2}$ (3). Then $\mathbf{WW_2} = \mathbf{R_1 S R_2} = \mathbf{A}$, where $\mathbf{A}$ is a general affine matrix.

WAT is not invertible. Clearly,

$$\mathbf{W}^{-1} = (\mathbf{R_1} \cdot \mathbf{S})^{-1} = \mathbf{S}^{-1} \cdot \mathbf{R_1}^{-1}. \quad (7)$$

The inverse of a rotation matrix is a matrix of an opposite rotation and the inverse of a non-uniform scaling matrix is again a non-uniform scaling matrix with inverse elements. However, rotation matrix and non-uniform scaling matrix generally do not commute, so $\mathbf{W}^{-1}$ does not have a form of the WAT matrix. (It should be noted that if only a *uniform* scaling was considered, then the commutativity would be guaranteed and both $\mathbf{W}$ and $\mathbf{W}^{-1}$ reduce to similarity transforms.)

Practical consequences for image registration are the following. When looking for the WAT mapping, we always have to specify the *direction* of the mapping. The WAT model is valid (at most) in one direction only (i.e., either from the reference image to the sensed one or vice versa) because it is not invertible. This is inconvenient when checking the *registration consistency*. In the consistency check the sensed image is first registered to the reference one, then the images are swapped and registered in the opposite direction. The registration errors are measured in both cases and the agreement between them is often used as an indicator of correctness of the whole registration process.

The closure property is required in group-wise and cascade registration. A typical example is a time series in which each frame is registered to the preceding one. Then the registration parameters between any two frames of the series are easily obtained by a composition of intermediate elementary registration parameters. This

[1] There exist also groups of non-linear transform models, such as projective group, but they are irrelevant to the topic of this paper.

approach can be used, among others, to validate the registration method – the composite parameters should be the same as those obtained by direct registration of the two respective frames. The composition is, however, possible only if the transformations are closed, which is not the case of WAT.

Although WAT does not constitute an affine subgroup, the affine group contains several other subgroups. We list those which are relevant to image registration.

- *Parameter-free subgroup* is only one – trivial subgroup containing the identity transform only.
- *One-parameter subgroups*
  - rotation around the origin,
  - uniform scaling,
  - translation in a given direction,
  - stretching (a special case of non-uniform scaling in which $s_2 = 1/s_1$),
  - horizontal/vertical skewing. Horizontal skewing is a transform with

$$\mathbf{A} = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}, \quad a \neq 0. \tag{8}$$

- *Two-parameter subgroups*
  - general translation,
  - rotation around the origin and uniform scaling,
  - non-uniform scaling.
- *Three-parameter subgroups* – general translation and rotation around the origin (i.e., rotation around an arbitrary point). This subgroup is called *rigid-body transform* group.
  - translation and uniform scaling,
  - translation and stretching.
- *Four-parameter subgroups*
  - translation and non-uniform scaling,
  - translation, rotation and uniform scaling (*similarity* group).
- *Five-parameter subgroups*
  - *area-preserving* affine group consists of the area-preserving affine transforms (APAT), i.e., affine transforms constrained by $\det \mathbf{A} = 1$.
  - *area-preserving* affine group with a *flip* consists of affine transforms constrained by $|\det \mathbf{A}| = 1$. This is a slight extension of the above group which allows also turning the sensed image "inside out" (which may be irrelevant in many applications such as in remote sensing and robot vision).

Hence, if there is a need for a five-parameter transform model and the area-preserving assumption is justified, we should use the area-preserving affine transform rather than the WAT.

4. Registration by means of the area-preserving affine transform (APAT)

Image registration by the APAT is justified whenever we know or assume that the area of the objects should be the same in the reference and sensed images. To calculate the transform parameters $(a_1, \ldots, a_6)$, two control points are insufficient while three and more CP's yield an over-determined system, the least-square solution of which leads to minimization of

$$E = \sum_{i=1}^{N} \|\mathbf{x}_i' - (\mathbf{A}\mathbf{x}_i + \mathbf{t})\|^2, \tag{9}$$

subject to the hard constraint of $\det \mathbf{A} = 1$ ($\mathbf{x}_i$ and $\mathbf{x}_i', i = 1, \ldots, N$ are the CP's in the reference and sensed images). There is no closed-form solution for the parameters; the use of Lagrange multipliers leads to a non-linear system of equations which must be resolved iteratively.

Fig. 2. The reference satellite image with the control points.

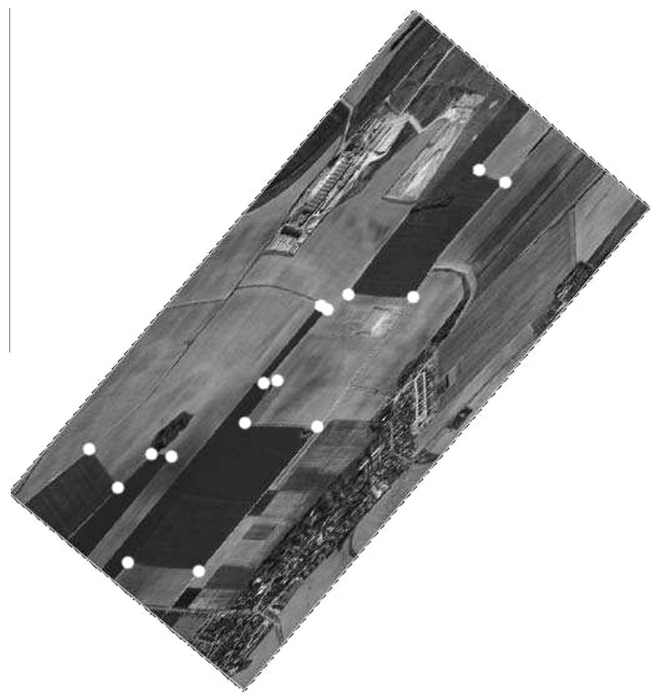

Fig. 3. The first sensed image was artificially created from the reference image by applying the WAT transform.

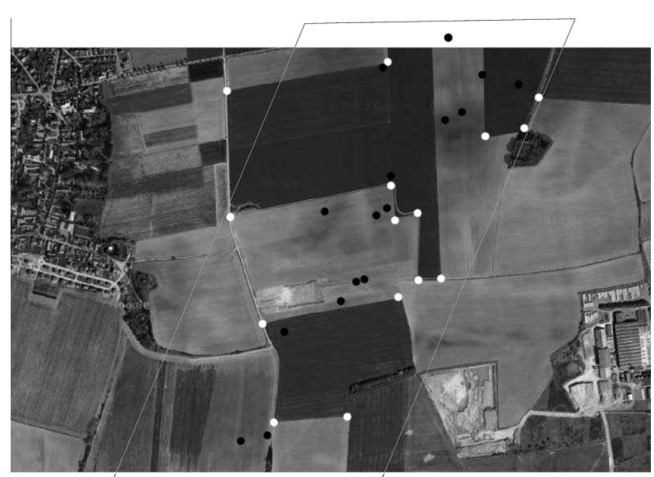

Fig. 4. The registration result. The sensed image was registered by the WAT model to the reference. The result (white frame) is overlayed over the reference. Note the inaccuracies at the control points.

Fig. 5. The second sensed image was created from the reference by applying two consecutive WAT transforms.

More flexible strategy is to relax the constraint $\det \mathbf{A} = 1$ and make it "soft" by incorporating the constraint into the functional which is minimized. Hence, we minimize

$$E_\lambda = \sum_{i=1}^{N} \|\mathbf{x}_i' - (\mathbf{A}\mathbf{x}_i + \mathbf{t})\|^2 + \lambda(\det \mathbf{A} - 1)^2, \tag{10}$$

where $\lambda > 0$ is a trade-off coefficient defined by the user which express the weight of the constraint. The minimization of E_λ is unconstrained, so it leads to the system of equations

$$\frac{\partial E_\lambda}{\partial a_k} = 0 \quad k = 1, \ldots, 6. \tag{11}$$

This system is unfortunately again non-linear and requires an iterative solution. Minimization of E_λ (10) instead of E (9) offers more flexibility and is better namely in situations where the area-preserving constraint holds only approximately. For $\lambda \to \infty$ both solutions are the same.

The APAT is not "always better" than the WAT. If the area-preserving assumption was significantly violated, the APAT registration would fail. However, thanks to its invertibility and closure property, APAT is applicable in numerous cases.

5. Illustrative experiment

To illustrate the above theory, we present a simple experiment for the readers' convenience. It clearly shows the differences between the two methods.

We transformed the reference image in Fig. 2 by a WAT transform (see Fig. 3)) to create the first sensed image. We intentionally choose such area-preserving transforms which include non-uniform scaling. We tried to register this sensed image to the reference by means of 16 control points (since the transform model is known exactly, there was no problem to find the correspondences). The transform coefficients were calculated by a least-square fit as suggested in Li and Zhang (2012). The result is highly inaccurate (see Fig. 4) and and far from the ground truth, the RMSE at the control points is 389 pixels. This is because the WAT model is not invertible; if we repeat this experiment using general affine transform and the APAT, both yield perfect results with subpixel accuracy.

We applied another WAT transform to the first sensed image and created the second sensed image (see Fig. 5), which was then registered by a WAT model to the reference. As can be seen from Fig. 6, the accuracy is unsatisfactory (RMSE = 341 pixels). Then we changed the order of registration and tried to register the reference image to the second sensed image by WAT model. The accuracy of the result (see Fig. 7) is only slightly better than in the previous case (RMSE = 205 pixels). Since it is not closed, the WAT model is inappropriate for registration in this case, even if the sensed image was generated by a consecutive exact WAT-warping and we tried both registration directions. We repeated the registration trial with the APAT mapping model. The mean-square error at the control points was less than one pixel

Fig. 6. The registration result. The second sensed image was registered by the WAT model to the reference. The result (white frame) is overlayed over the reference. Note the inaccuracies at the control points.

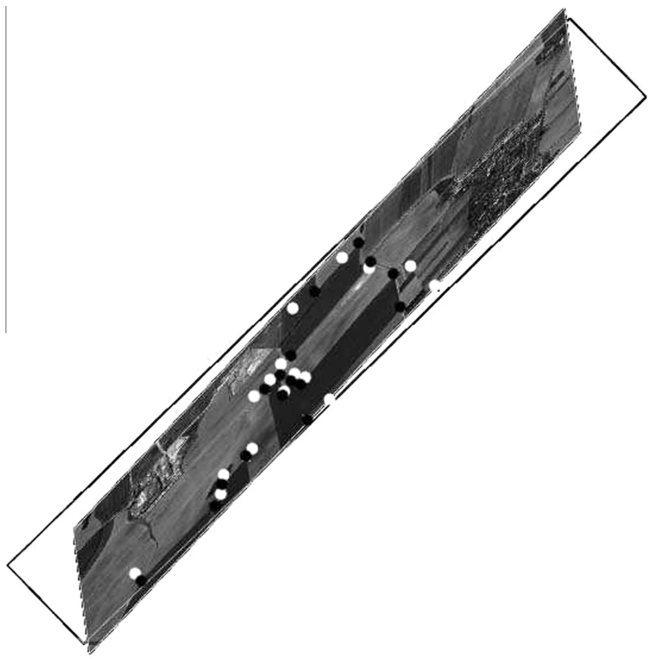

Fig. 7. The registration in the reverse order. The reference image was registered by the WAT model to the second sensed image. The accuracy is only slightly better than in the previous case.

for both registration directions. The same result was achieved by an unconstrained affine transform, which would work with the same accuracy even if the warping is not area-preserving.

6. Conclusion

The contribution of this paper is twofold. First, we exposed certain difficulties of the WAT transform which were ignored in Li and Zhang (2012) and which would have a negative impact if the WAT transform was used for image registration. Second, we show that there exists another five-parametric transform which forms a subgroup of the affine group and thus can be used as a mapping model for image registration purposes.

Acknowledgement

This work has been supported by the Grant No. P103/11/1552 of the Czech Science Foundation.

References

Li, D., Zhang, Y., 2012. A novel approach for the registration of weak affine images. Pattern Recognition Letters 33, 1647–1655.

Suk, T., Flusser, J., 2005. Affine normalization of symmetric objects. Proc. Advanced Concepts for Intelligent Vision Systems ACIVS 2005, LNCS vol. 3708. Springer, pp. 100–107.

Zitova, B., Flusser, J., 2003. Image registration methods: a survey. Image and Vision Computing 21 (11), 977–1000.