A Robust Outlier Elimination Approach for Multimodal Retina Image Registration

Ee Ping Ong[1], Jimmy Addison Lee[1], Jun Cheng[1], Guozhen Xu[1], Beng Hai Lee[1], Augustinus Laude[2], Stephen Teoh[2], Tock Han Lim[2], Damon W.K. Wong[1], and Jiang Liu[1]

[1]Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore
{epong,jalee,jcheng,benghai,xug,wkwong,jliu}@i2r.a-star.edu.sg
[2]National Healthcare Group (NHG), Eye Institute, Tan Tock Seng Hospital, Singapore
{laude_augustinus,stephen_teoh,tock_han_lim}@ttsh.com.sg

Abstract. This paper presents a robust outlier elimination approach for multimodal retina image registration application. Our proposed scheme is based on the Scale-Invariant Feature Transform (SIFT) feature extraction and Partial Intensity Invariant Feature Descriptors (PIIFD), and we combined with a novel outlier elimination approach to robustly eliminate incorrect putative matches to achieve better registration results. Our proposed approach, which we will henceforth refer to as the residual-scaled-weighted Least Trimmed Squares (RSW-LTS) method, has been designed to enforce an affine transformation geometric constraint to solve the problem of image registration when there is very high percentage of incorrect matches in putatively matched feature points. Our experiments on registration of fundus-fluorescein angiographic image pairs show that our proposed scheme significantly outperforms the Harris-PIIFD scheme. We also show that our proposed RSW-LTS approach outperforms other outlier elimination approaches such as RANSAC (RANdom SAmple Consensus) and MSAC (M-estimator SAmple and Consensus).

1 Introduction

Image registration is the process of aligning an image to another image taken from the same scene/object but in different situations [7, 10, 15]. Image registration is an important prior processing step before other processes such as image mosaicking [2], image fusion [7], and retina verification etc. can be carried out. Image registration approaches can be classified into two different types: unimodal and multimodal [7, 15]. Multimodal image registration approaches are designed for registering pairs of images from different imaging modalities, such as attempting to register a retina color fundus image with a retina fluorescein angiographic (FA) image. The purpose of such multimodal retina image registration is to assist eye doctors to automatically register fundus and FA images captured from the retina of the same person to enable the doctor to more easily diagnose eye problems and diseases. The registered fundus-FA image pair and their fusion can aid the doctor in planning treatment and surgery for their patients.

On the other hand, unimodal image registration approaches are designed for registering pairs of images from the same imaging modality. Generally, unimodal image registration approaches do not work on multimodal image registration.

One of the better known multimodal image registration approach in recent year is the Harris-partial intensity invariant feature descriptor (Harris-PIIFD) approach proposed by Chen et al. [3]. In [3], they proposed a scheme to register multimodal images by extracting corner points using the Harris-Stephen corner detector, estimate the main orientation of each corner point, followed by computing a partial intensity invariant feature descriptor (PIIFD) for each point. Then, they match the PIIFDs using a nearest-neighbor criterion with a pre-determined threshold to obtain the putative matches. Thereafter, they attempt to remove any incorrect matches (i.e. outliers) by applying 2 different pre-determined thresholds to the 2 criteria: the difference in main orientations and the ratio of geometric distances between corner points. Then, they try to refine the locations of the matches. Thereafter, they apply a particular transformation mode to register the pair of multimodal images to be registered. Chen et al. [3] showed that their proposed Harris-PIIFD approach performs better than SIFT [8] and GDB-ICP [12, 13] algorithms for registering multimodal image pairs.

In [6], the authors proposed the use of uniform robust scale invariant feature transform (UR-SIFT) features (instead of Harris corner feature points as used by [3]) together with PIIFD for improved multimodal image registration. Incorrect matches (i.e. outliers) are eliminated by checking each matched pair against the global transformation and those which have highest geometric errors are removed one by one until achieving a root mean square error less than a certain threshold.

In both approaches proposed in [3] and [6], the outliers elimination step appears weak and are not robust enough when there are very high percentage of incorrect matches present in the putatively matched feature points. An alternative robust outlier elimination approach that is widely used in the computer vision community is the RANdom SAmple Consensus (RANSAC) algorithm [5]. The goal of RANSAC is to find a model that maximizes the number of data points with error smaller than a user defined threshold t. One of the problems with the RANSAC approach is that in many real-life problems, it might not be possible to know the error bound t for inlier data beforehand. As pointed out in [14], the choice of the threshold is a sensitive parameter and can affect the performance dramatically. In [4], the authors conducted a performance evaluation of the various robust outlier elimination methods in the RANSAC family (such as RANSAC, MSAC, MLESAC, LO-RANSAC, R-RANSAC, MAPSAC, AMLESAC, GASAC, u-MLESAC, and pbM-estimator) for a range of outlier ratiso from 10% to 70%. They found that many methods do not perform well when the outlier ratio was increased to 70% despite the fact that various thresholds associated with the respective methods were tuned and the authors only report the best results obtained. The authors [4] showed that at outlier ratio of 70%, MSAC (M-estimator SAmple and Consensus) [14] and pbM-estimator give the best results.

Our contributions are as follow: We propose a robust outlier elimination approach, which hereby we will refer to as the residual-scaled-weighted Least Trimmed Squares (RSW-LTS). We will show how we can adapt the Least Trimmed Squares estimator (using our proposed formulation of weights estimate and iteratively re-weighted least squares procedure) to apply an affine transformation geometric constraint for robust pruning of putatively matched feature points (to remove the wrong matches or outliers) and to estimate the affine transformation model's parameters. We show that we are able to achieve the goal of obtaining improved performance over the original Harris-PIIFD multimodal scheme in [3], as well as improved performance over other outlier elimination approaches such as RANSAC [5] and MSAC [4, 14].

2 Proposed Method

Our proposed scheme operates as follow:

1. Extract salient feature points using the scale-invariant feature transform (SIFT) algorithm [8].
2. Compute the partial intensity invariant feature descriptor (PIIFD) [3] for each salient feature point (using the method in [3]).
3. Putative matching of the PIIFDs using a nearest-neighbor criterion (following the method described in [3]).
4. Outlier elimination and affine transformation parameter estimation using our proposed approach.
5. Perform image registration using the estimated transformation parameters.

The details of our proposed outlier elimination approach are as described below.

2.1 Outlier Elimination and Transformation Parameter Estimation

The putative matches obtained in earlier stage may have many wrong matches (or outliers), depending on the complexity and difficulty of the matching due to the underlying differences between the pair of images to be registered. As such, a robust outlier elimination step is required. Here, we proposed the residual-scaled-weighted Least Trimmed Squares (RSW-LTS) estimator with affine transformation to perform outlier elimination. In essence, we are trying to apply an affine transformation geometric constraint on the putative matches and remove those matched points that do not satisfy the same geometric constraint. The robust estimator we proposed here is basically a Least Trimmed Squares (LTS) estimator with residual-scaled-weights and followed by an iteratively re-weighted Least Squares estimation process.

The original Least Trimmed Squares (LTS) estimate [11] is given by:

$$\hat{\boldsymbol{\Theta}} = \arg\min_{\boldsymbol{\Theta}} \sum_{i=1}^{h} \hat{r}_i(\mathbf{x}, \boldsymbol{\Theta})^2 \quad (1)$$

where Θ is the set of model parameters, $\mathbf{x} \in R$, R being the neighborhood used for the transformation computation, $\hat{r}_1 <= \ldots <= \hat{r}_n$ are the ordered squared residuals, n is the total number of data in the dataset, and h denotes the h^{th} ranked residual above which the rest of the residuals representing (n-h) will be trimmed (i.e. removed from consideration in the estimation of the model parameters). Rousseeuw and Leroy [11] states that h should be more than $[n/2]+1$, giving the LTS estimator a maximum outlier breakdown point of 50%.

The residual-scaled-weighted Least Trimmed Squares affine transformation estimation method that we proposed here has the advantage that it does not need a pre-defined threshold t for inliers (in contrast to the RANSAC method). Without the need for a pre-defined t, the proposed method is very flexible and is able to handle different data, and hence has clear advantage over the standard RANSAC method. In addition, contrary to the value of h advocated in [11], we found that we are able to obtain good results with our proposed scale estimate, residual-scaled weights, and iteratively re-weighted least-squares procedure for our adapted LTS method even when we have chosen a value of h less than $[n/2]+1$. This means that our proposed RSW-LTS with the associated scale estimate and weights computation coupled with the iteratively re-weighted least squares computation can give an outlier rejection ratio of more than 50%. By using a value of h equals to $n/10$, our experiments show that we are able to achieve up to about 90% of outlier rejection. The details of the residual-scaled-weighted Least Trimmed Squares affine transformation estimator are as follow.

An affine transformation model that describes the affine transformation of a point $\mathbf{x} = (x, y)^T$ to $\mathbf{u} = (u, v)^T$, in the x and y directions, can be defined as:

$$\mathbf{u}(\mathbf{x}, \boldsymbol{\Theta}) = \mathbf{A}\mathbf{x} + \mathbf{b} \tag{2}$$

where $\mathbf{A} = \begin{pmatrix} a_0 & a_1 \\ a_2 & a_3 \end{pmatrix}$ and $\mathbf{b} = (b_0, b_1)^T$, and we can denote the transformation parameters as: $\boldsymbol{\Theta} = (b_o, a_0, a_1, b_1, a_2, a_3)^T$. Let the residual in modelling the transformation (with transformation parameters $\boldsymbol{\Theta}$) at a point $\mathbf{x}$ be $r(\mathbf{x}, \boldsymbol{\Theta})$, where:

$$r(\mathbf{x}, \boldsymbol{\Theta}) = \left\| \mathbf{u}(\mathbf{x}, \boldsymbol{\Theta}) - \mathbf{u} \right\| \tag{3}$$

Given a data set of n observations, the algorithm repeatedly draws m sub-samples each of p different observations from the data set using a Monte-Carlo type technique, where p is the number of parameters in the model (p = 3 for the affine transformation model). For each sub-sample, indexed by J, $1 \le J \le m$, the corresponding parameters (denoted by $\hat{\boldsymbol{\Theta}}_J$) are estimated from the p observations. In addition, the sum of the least trimmed squared residuals, denoted by M_J, is also determined, where:

$$M_J = \sum_{i=1}^{h} \hat{r}_i(\mathbf{x}, \hat{\boldsymbol{\Theta}}_J)^2 \tag{4}$$

The LTS solution is the $\hat{\Theta}$ for which the corresponding M_J is the minimum among all the m different M_Js, i.e.:

$$\hat{\boldsymbol{\Theta}} = \underset{\boldsymbol{\Theta}_J}{\operatorname{argmin}} \sum_{i=1}^{h} \hat{r}_i(\mathbf{x}, \boldsymbol{\Theta}_J)^2 \tag{5}$$

The minimum number of sub-samples, m, required is given by:

$$m = \frac{\log(1-P)}{\log[1-(1-\varepsilon)^p]} \tag{6}$$

where P is the probability that at least one of the m sub-samples consists of p good observations, and ε is the fraction of outliers that may be present in the data.

We proposed the following to estimate the initial scale estimate of RSW-LTS:

$$\sigma^0 = c.\frac{1}{h}\sum_{i=1}^{h}\hat{r}_i(\mathbf{x},\hat{\mathbf{\Theta}}) \tag{7}$$

where $h = n/10$ (so as to be able to handle up to 90% of outliers), and c is a constant and is merely a correction factor used to achieve consistency at Gaussian error distributions. The initial scale estimate is used to determine an initial weight $w(\mathbf{x})$ for each observation using the following formulation:

$$w(\mathbf{x}) = \begin{cases} \left(e^{-\left(r(\mathbf{x},\hat{\mathbf{\Theta}})/\sigma^0\right)^2/2}\right) \Big/ \sqrt{2\pi}\sigma^0 & \text{if } \left|r(\mathbf{x},\hat{\mathbf{\Theta}})/\sigma^0\right| \le d \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where d is a constant and determines the percentage of Gaussian inliers that may be incorrectly rejected. We have used a value of 1.96 for d so that Gaussian inliers are only incorrectly rejected 5% of the time. Note that in our case, the weights used are scaled using the scale estimate with respect to the residual values. The reason we used residual-scaled weights is because doing so will remove the data dependency and enable the method to handle different data (in contrast to, for example, RANSAC which uses a fixed pre-defined threshold to filter the residuals, and resulted in needing to use different threshold for different data in order to get optimal results).

Thereafter, we proposed an iterative re-weighted least squares procedure to iteratively estimate the transformation parameters $\hat{\Theta}^j$ by solving:

$$\hat{\Theta}^j = \underset{\Theta}{\operatorname{argmin}} \sum_{\mathbf{x}} w^j(\mathbf{x}) r^j(\mathbf{x},\Theta^{j-1})^2 \tag{9}$$

We performed the iterative re-weighted least squares computation (according to equation (9)) for 10 iterations (for j = 1, ..., 10) using updated σ^j (according to equation (7) by replacing σ^0 by σ^j) and updated w^j ($\mathbf{x}$) (according to equation (8) and with σ^0 replaced by σ^j) for each of the j^{th} iteration. In our experiments, we found that σ^j will usually converge after 4-6 iterations.

3 Results

Our proposed method has been evaluated on a dataset collected by clinicians from a Hospital. It contains a total of 200 retina images (100 pairs of corresponding color fundus and fluorescein angiographic (FA) images). These are real-life fundus-FA image pairs of patients showing symptoms of severe macula edema and staphyloma that necessitate retinal photocoagulation or photodynamic therapy. These fundus-FA image pairs are referred to by doctors in the pre-treatment stage to guide them in their planning of treatments for their patients. This dataset is very challenging because they are retina images of real-life diseased eyes and is a particular dataset considered by the doctors as most challenging ones compared to other retinal abnormalities.

To illustrate the operations of our proposed retina image registration scheme, we show the original color fundus and FA images in Figure 1, while Figure 2 shows the feature points extracted from both the fundus and FA images using the Scale-Invariant

Feature Transform (SIFT) algorithm. Figure 3 shows the results of the putative feature point matching step using the Partial Intensity Invariant Feature Descriptors (PIIFD), while Figure 4 shows the results of pruned matched feature points after outlier elimination using our proposed RSW-LTS approach. From this example, you can see that there are a total of 143 putatively matched feature points (see Figure 3). However, a very large percentage of these putative matches are incorrect. After applying our proposed RSW-LTS outlier elimination approach, we obtained only 10 matched feature points which are all correctly matched (see Figure 4). In this example, the outlier rejection ratio is about 93%. Theoretically, our RSW-LTS approach has been designed to handle up to 90% of outliers. However, in practice, our proposed approach has been shown to work for slightly more than 90% of outliers, as in this example (due to the combinations of scale estimate, residual-scaled weights and the iteratively reweighted least squares procedure). (Note that this example shows the image pair that the Harris-PIIFD scheme [3] is unable to register successfully, and both the RANSAC [5] and MSAC [14] methods also failed.)

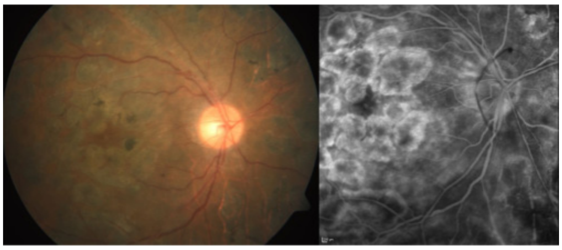

Fig. 1. riginal fundus-FA image pair.

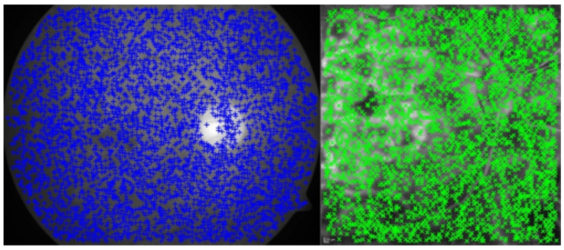

Fig. 2. Extracted SIFT feature points.

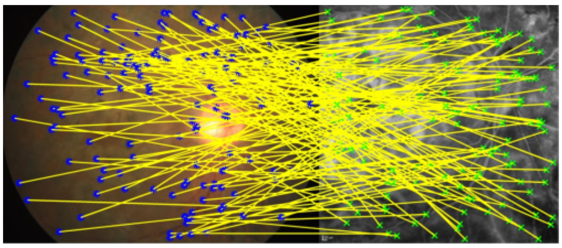

Fig. 3. Results of PIIFD putative matching.

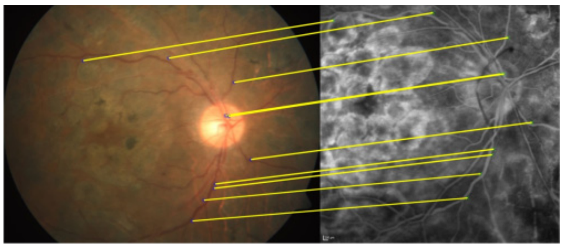

Fig. 4. Pruned matches using RSW-LTS.

In order to objectively evaluate the fundus-FA image registration results, we have adopted the root-mean-squared error (RMSE) [6, 9], the median error (MEE) [3, 6], and the maximal error (MAE) [3, 6] as the objective error measurements. These objective error measures are computed from the ground truth data generated for all the fundus-FA image pairs. We obtained the ground truth data by manually selecting 10 corresponding points that are distributed uniformly in each of the corresponding fundus and FA image pair. The ground truth and their localization are then double-checked by another group of team members to verify the correctness of the manually marked corresponding points. The process to generate the ground truth data is tedious and time consuming but necessary so that we are able to compute the objective error measures for different image registration techniques. Following [6], we consider an image pair to have been successfully registered when its RMSE is below 5 pixels, since RMSE less than 5 pixels have been found to be acceptable for clinical purposes [9]. We adopted the definition of the “success rate” to be the ratio of the number of successfully registered image pairs to the total number of image pairs being registered [6, 12]. The average objective error measures are then computed only for those image pairs with successful registration.

Table 1 shows the results of Harris-PIIFD scheme [3] and our proposed scheme for successful fundus-FA image registration. It can be seen that our proposed scheme with 83% successful registration rate significantly outperforms the original Harris-PIIFD scheme [3] (which only has a 22% successful registration rate). Also, our proposed scheme gives lower objective measurement errors (in terms of average RMSE, MAE, and MEE) compared to the Harris-PIIFD scheme. We found that we are unable to obtain 100% successful registration rate because some of these images are of very poor visual quality and/or correspondence for the purpose of image registration.

In order to verify that our RSW-LTS approach contributes to the significant improvements in successful registration rate and lower objective measurement errors (and not because of the SIFT+PIIFD combination), we had conducted several other experiments where we replaced our RSW-LTS algorithm with the well-known robust estimator, RANSAC [5], while keeping all other processes the same as in our proposed scheme with RSW-LTS. (We used publicly available RANSAC implementation to generate our results here). The number of random sub-samples used for RANSAC has been set to very large in all RANSAC experiments in order to reduce the effect of error due to sampling (We used a value of 9000 in all our experiments, thus the confidence of finding a good sub-sample and parameter estimate here is about 99.9%). Note that the main parameters that control the performance of the RANSAC method are the number of random sub-samples used and the threshold applied (which determines the error tolerance allowed in considering whether a matched point is an inlier). We experimented with numerous well-spaced threshold values for RANSAC and we only present the best RANSAC result (obtained with a specific threshold value) in Table 1. From Table 1, it can be seen that the RANSAC approach only has a successful registration rate of 32%. We analyzed the results and found that in many cases, the percentage of the putative feature point matches that is wrong is quite large, and there are significant number of cases where the outlier ratio range from 70% to more than 90%. This could be the reason why RANSAC did not perform well in our case here, despite the tuning of the parameters.

Table 1. Comparison of results.

Method	Success rate (%)	Average RMSE	Average MAE	Average MEE
Harris-PIIFD scheme [3]	22%	3.40	8.22	3.93
Our proposed scheme (with RSW-LTS)	83%	3.08	7.80	3.50
Our scheme but with LTS replaced by RANSAC	32%	3.23	8.12	3.64
Our scheme but with LTS replaced by MSAC	34%	3.16	7.81	3.71

Since MSAC [14] has been reported in [4] to be one of the two best methods at outlier ratio of 70%, we also presented the image registration results obtained by replacing the RSW-LTS with MSAC. (We used publicly available MSAC implementation to generate the results here). The results of the approach with the RSW-LTS replaced with MSAC (and with a threshold that gives the best results) are shown in Table 1, where it only gives a successful registration rate of 34%. From the above, we found that our proposed RSW-LTS approach when applied for fundus-FA image registration gives the best results (in terms of significantly higher successful registration rate and lower objective measurement errors) when compared to using RANSAC [5], MSAC [14], and the Harris-PIIFD scheme [3]. The computational time for Harris-PIIFD [3] is about 3.9s, our proposed scheme is 34.0s, using RANSAC is 36.1s, and using MSAC is 52.2s (when running all methods on the same Intel i7-4470 3.40GHz PC with 8GB RAM in Matlab). However, as our proposed approach is for the purpose of multimodal retina image registration application that doctors may use during pre-planning stage for planning of treatment and surgery for their patients, real-time computation is not necessary. Hence, our study here is primarily focused on the robustness and increasing the accuracy and successful registration rate of the multimodal image registration application.

4 Conclusions

In this paper, we present a robust outlier elimination approach for multimodal retina fundus-fluorescein angiographic image registration. Our main contribution here is a novel outlier elimination method, the residual-scaled-weighted Least Trimmed Squares (RSW-LTS), to robustly eliminate incorrectly matched putative feature point matches. We showed that our proposed scheme provides significant improvement over the original Harris-PIIFD scheme [3]. The improvement in performance is mainly due to our outlier elimination method. We also showed that our proposed RSW-LTS approach outperforms other outlier elimination approaches such as RANSAC (RANdom SAmple Consensus) and MSAC (M-estimator SAmple and Consensus).

Acknowledgement. The authors are grateful to Jian Chen for providing his codes that implement [3].

References

1. Brown, M., Lowe, D.G.: Recognising panoramas. In: Int. Conf. Comp. Vis, pp. 1218–1225 (2003)
2. Cattin, P.C., Bay, H., Van Gool, L., Székely, G.: Retina mosaicing using local features. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 185–192. Springer, Heidelberg (2006)
3. Chen, J., Tian, J., Lee, N., Zheng, J., Smith, R.T., Laine, A.F.: A Partial Intensity Invariant Feature Descriptor for Multimodal Retinal Image Registration. IEEE T. on Biomedical Engineering 57, 1707–1718 (2010)

4. Choi, S., Kim, T., Yu, W.: Performance evaluation of RANSAC family. In: Proc. British Machine Conference, pp. 81.1–81.12 (2009)
5. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. of the ACM 24(6), 381–395 (1981)
6. Ghassabi, Z., Shanbehzadeh, J., Sedaghat, A., Fatemizadeh, E.: An efficient approach for robust multimodal retinal image registration based on UR-SIFT features and PIIFD descriptors. EURASIP J. on Image and Video Processing, 25 (2013)
7. Laliberte, F., Gagnon, L., Sheng, Y.: Registration and Fusion of Retina Images – An Evaluation Study. IEEE T. Medical Imaging 22(5), 661–673 (2003)
8. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. Int. J. of Computer Vision 60(2), 91–110 (2004)
9. Matsopoulos, G.K., Asvestas, P.A., Mouravliansky, N.A., Delibasis, K.K.: Multimodal registration of retinal images using self organizing maps. IEEE T. on Med. Imaging 23(12), 1557–1563 (2004)
10. Ritter, N., Owens, R., Cooper, J., Eikelboom, R., van Saarloos, P.P.: Registration of stereo and temporal images of the retina. IEEE T. Medical Imaging 18, 404–418 (1999)
11. Rousseeuw, P.J., Leroy, A.M.: Robust regression and outlier detection. John Wiley & Sons, New York (1987)
12. Stewart, C.V., Tsai, C.L., Roysam, B.: The dual-bootstrap iterative closest point algorithm with application to retinal image registration. IEEE T. Med. Imaging 22(11), 1379–1394 (2003)
13. Stewart, C.V., Yang, G.: The generalized dual bootstrap-ICP executable (2008). `http://www.vision.cs.rpi.edu/download.html`
14. Torr, P.H.S., Zisserman, A.: MLESAC: A new robust estimator with application to estimating image geometry. J. Computer Vis. & Image Understanding 78(1), 138–156 (2000)
15. Zitova, B., Flusser, J.: Image registration methods - A survey. IVC 21, 977–1000 (2003)