PERFORMANCE EVALUATION FOR SCENE MATCHING ALGORITHMS BY SVM

Zhaohui Yang [a, b,] *, Yingying Chen [a], Shaoming Zhang [a]

[a] The Research Center of Remote Sensing and Geomatic, Tongji University, Shanghai 200092, China - yzhac@163.com

[b] School of Environmental Science and Engineering, Suzhou University of Science and Technology, Suzhou 215011, China - yzhmilan@gmail.com

Commission VII, WG VII/4

KEY WORDS: Scene matching, Classification, Support vector machine, Performance evaluation, Measure descriptor, Training

ABSTRACT:

Scene matching is the process of locating a region of an image with the corresponding region of another image where both image regions represent the same scene. Although a lot of algorithms have appeared on scene matching, performance analysis is usually based on simple statistic experiment and performed simply and visually, and little attention has been given to evaluate performance of different algorithms. In order to choose suitable algorithms and improve the performances of the algorithms, we present a novel performance evaluation method for scene matching algorithms based on support vector machine (SVM), which can partly show interact-effect of numerous similarity measure factors and find a dependency link between two correlative images. The method is described with a three-step procedure. Firstly we build samples data set using similarity measure descriptors of image pairs. Then decision function is obtained through training and testing process with input of samples data. Finally, we adopt result of SVM classification to evaluate two classical algorithms: normalized cross-correlation algorithm and Canny-based edge extraction algorithm. The experimental results show that this method holds the capability of automatic decision ability for performance evaluation and high ratio of correct prediction.

1. INTRODUCTION

Scene matching refers to the process of locating a region in one image with the corresponding region in another image where both image regions represent the same scene. The two images are often taken under different time, different sensor pose geometry, or taken with different types of sensor (Sjahputera, 2004). The main goal in scene matching is to assess the degree of similarity and find a dependency link between two correlative images. Now scene matching is a widely used technology in real-time applications such as flight navigation and missile guidance.

Some classical algorithms have been developed lies in the variety of sensor styles and variety of image characteristics. In order to choose suitable algorithms and improve the performances of the algorithms, some researchers have studied performance evaluation for matching algorithms (Moigne, et al., 1998; Coutre, et al., 2000), which is essential for the successful creation and interpretation the criterion of test data of matching algorithm. However, defining inter-comparison criteria is a difficult task, since each algorithm should take into account not only the geometric distortion between the images but also radiometric deformations and noise corruption and application-dependent images characteristics. Therefore, an accurate mathematical model that is able to incorporate all similarity measure descriptors and can predict the outcome of matching algorithms is not feasible. Currently, most of performance evaluation methods are based on simple statistic experiment, which cannot deal with interact effect of numerous similarity measure factors. Furthermore, the statistic methods are useful for large-scale train data set, but they do not seem suited for the problem only with a few of examples available.

To overcome above drawback, we introduce the support vector machine (SVM) to performance evaluation for scene matching algorithms. Firstly we build a low number of sample data set using similarity measure descriptors of pairs of reference and sensed images. Then decision function is trained through training and testing with input of samples data. Finally, we use SVM classification output to evaluate classical algorithms. Thus we can obtain an objective performance evaluation for scene matching algorithms at same condition without any priori knowledge.

2. SUPPORT VECTOR MACHINE

Support vector machine (SVM) lies in strong connection to the statistical learning theory, where it implements the structural risk minimization for solving two class classification problems (Vapnik, 1998). SVM transforms the input into a high-dimensional space using a nonlinear mapping. To overcome the increased computational complexity and over-fitting problems caused by the transformation, SVM constructs a maximum margin hyper-plane and support vectors. The maximum margin hyper-plane is set as far away as possible between classes, and the support vectors are the instances that are closest to the maximum margin hyper-plane.

Given a training set of instance-label pairs $\{y_i, x_i\}_{i=1}^{N}$, with each input $x_i \in R^n$ and the output label $y_i \in R$, the support

* Corresponding author

vector method aims at building a decision function for classification as follow.

$$y(x) = sign\left[\sum_{i=1}^{N} \alpha_i y_i \langle \varphi(x_1), \varphi(x_2) \rangle + b\right] \tag{1}$$

where α_i = positive real constants
b = real constant
φ = nonlinear mapping
$\langle \cdot , \cdot \rangle$ = inner product

Using the kernel function $K(x_i, x_j)$ instead of the inner product, the low-dimensional input could be mapped into the high-dimensional space and Eq.(1) could be replaced as follow.

$$y(x) = sign\left[\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b\right] \tag{2}$$

the classifier is constructed as

$$y_i = \left[w^T \varphi(x_i) + b \right] \geq 1 \qquad i = 1, 2, \ldots N \tag{3}$$

When the training set is not separable, the SVM algorithm tries to minimize $\|w\|$ under the condition of separating the data with a minimum number of errors. This is improved by slack variable ζ_i and penalty parameter c. Thus SVM requires the solution of the following optimization problem:

$$\begin{cases} \min\limits_{w,b,\zeta} \quad \frac{1}{2} w^T w + c \sum_{i=1}^{N} \zeta_i \\ subject \quad to: y_i = \left[w^T \varphi(x_i) + b \right] \geq 1 - \zeta_i \qquad i = 1, 2, \ldots N \\ \zeta_i \geq 0 \end{cases} \tag{4}$$

In least squares support vector machines (LS-SVM), least squares version is related to the cost function (Suykens and Vandewalle, 2000). The optimization problem is modified into

$$\begin{aligned} &\min \tau(w, \zeta_i) = \frac{1}{2} w^T w + \frac{c}{2} \sum_{i=1}^{N} \zeta_i^2 \\ &subject \quad to: \quad y_i \left[w^T \varphi(x_i) + b \right] = 1 - \zeta_i \qquad i = 1, 2, \ldots N \end{aligned} \tag{5}$$

the corresponding Lagrangian is

$$L(w, b, \zeta, \alpha) = \tau(w, \zeta) - \sum_{i=1}^{N} \alpha_i \left\{ y_i \left[w^T \varphi\left(x_i\right) + b \right] - 1 + \zeta_i \right\} \tag{6}$$

the optimality condition leads to the set of linear equation as

$$\begin{bmatrix} 0 & -Y^T \\ Y & ZZ^T + c^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \vec{1} \end{bmatrix} \tag{7}$$

where $z = \left[\varphi(x_1)^T y_1, \ldots, \varphi(x_N)^T y_N \right]$
$Y = \left[y_1, \ldots, y_N \right]$
$\vec{1} = [1, \ldots, 1]$
$\alpha = [\alpha_1, \ldots, \alpha_N]$

3. CLASSIC SCENE MATCHING ALGORITHMS

Classic scene matching methods can be divided into two major groups (Brown, 1992; Zitova and Flusser, 2003): area-based algorithms that used images pixel intensity values directly (Barnea and Silverman, 1972) and feature-based algorithms that use obvious features such as edges and points (Wong, 1980; Terefa and Harada, 2001). Recently, scene matching methods using simultaneously both area-based and feature-based approaches have started to appear.

Area-based algorithms have good capability of representing the image's intensity character under the low distortion and low greyscale difference condition. However, they are just a type of simple similarity measurement and are weak-robust to noise or great distortion. On the other hand, feature-based algorithms, which match features represent information on global level, are typically applied when the local structural information is more significant than the intensity information. This property makes feature-based methods suitable for situations when great illumination and distortion changes are expected or multi-sensor image analysis is needed.

Here, we choose two classical algorithms as research targets of performance evaluation. One is the normalized cross-correlation algorithm (NCC) which is representative of the area-based methods. NCC algorithm computes the measure of similarity for window pairs of images and searches its maximum as matching location. Despite high computational complexity, this method is still often in use, particularly thanks to its easy hardware implementation, which makes it suitable for real-time applications. Another is Canny-based edge extraction algorithm (CEE) which belongs to feature-based algorithms. The method eliminates background edges in images by making smooth filter firstly. Then using Canny edge detector, only salient edges are extracted by adjusting the threshold to minimize the weak edges. Finally, the couple of binary edge images are matched and the pairwise correspondence between reference image and sensed image is obtained using their spatial edge relations of features.

4. SIMILARITY MEASURE DESCRIPTORS

In order to evaluate performance of scene matching algorithms, we should extract measure descriptors (also called features or parameters) from reference and sensed image information. Once the measure descriptors have been achieved, they must be coded as a description vector which acts as the input of SVM. In our method, the measure descriptors can be classified in two groups: gray-based descriptors and edge-based descriptors. The former are directly obtained from image intensity and assess statistical characteristics, and the latter are gained from edge information of image and describe the unique structure features. These two level descriptors actually represent image's information from fine to coarse scale, and they are supplementary from each other.

There are lots of methods to gain the edge information, such as the edge detection algorithm, the local image gradients, etc. In this paper, we adopt Canny (Canny, 1986) operator to extract the edge information due to its excellent capability of accurate localization and responses to a single edge.

Here, we totally extract two gray-based descriptors (mean variance of image and image entropy) and two edge-based descriptors (edge density and geometric invariants).

4.1 Variance of image

The variance of image is the average squared deviation of all pixels from the sample mean. The variance of image, *var*, is computed using the equation

$$var = \frac{1}{M \times N} \sum_{i=0}^{M} \sum_{j=0}^{N} [I(i,j) - E]^2 \tag{8}$$

where $M \times N$ = image size
$I(i,j)$ = intensity in position of (i,j) of reference image
E = average of reference image intensity.

4.2 Image entropy

Image entropy is a quantity which is used to character the certain quality of an image. Low entropy images lack of detail information and high entropy images have a great deal of contrast information. Consequently high entropy images can show the more detail information of image. Image entropy is calculated with follow formula

$$H_f = -\sum_{i=1}^{M} \sum_{j=1}^{N} p_{ij} \log p_{ij} \tag{9}$$

$$p_{ij} = f(i,j) / \sum_{i=1}^{M} \sum_{j=1}^{N} f(i,j) \tag{10}$$

where p_{ij} = probability of difference intensity of (i,j)
$f(i,j)$ = gray value of pixel (i,j) in reference image

4.3 Edge density

Edge density can show the concentration of features in original image. We use edge density *ED* as descriptor, which is computed from binary Canny-edge image by

$$ED = \frac{Canny(I)}{M \times N} \tag{11}$$

where $Canny(I)$ = total number of edge points
$M \times N$ = image size

4.4 Geometric invariants

Another descriptor computed from extracted edges will be used: the geometric invariants. A family of seven invariants with respect to planar transformations is firstly obtained by Hu (Hu, 1962). Those invariants are invariants to rotation, scaling and translation, which can be regarded as nonlinear combinations of complex geometric moments:

$$c_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x+iy)^p (x-iy)^q f(x,y) dxdy \tag{12}$$

where $x,\ y$ = coordinates of the image $f(x,y)$
i = imaginary unit
$p+q$ = order of c_{pq}

Here we use modified invariants (Flusser, 2000), which have the following forms:

$$\begin{array}{ll}
\psi_1 = c_{11} & \psi_2 = c_{21}c_{12} \\
\psi_3 = re(c_{20}c_{12}^2) & \psi_4 = im(c_{20}c_{12}^2) \\
\psi_5 = re(c_{30}c_{12}^3) & \psi_6 = im(c_{30}c_{12}^3) \\
\psi_7 = c_{22} & \psi_8 = re(c_{31}c_{12}^2) \\
\psi_9 = im(c_{31}c_{12}^2) & \psi_{10} = re(c_{40}c_{12}^4) \\
\psi_{11} = im(c_{40}c_{12}^4) &
\end{array} \tag{13}$$

5. SVM CLASSIFICATION

5.1 Samples data set construction

The sample data set is obtained by combination of SPOT 5 panchromatic images, ETM+ band 8 panchromatic images and SAR images. SPOT 5 image patches regard as reference images, and ETM image patches and SAR image patches simulate as sensed images. In this way, each type of sensed image and corresponding reference image compose a pair of sample images. We process the scene matching between each pair of sample image using NCC algorithm and CEE algorithm respectively. According to the matching location offset, we classify two labels as right matching class (less than three pixels offset) and wrong matching class (great than three pixels offset) for each algorithm. That is traditional two-class problems in SVM classification.

We select some pairs of sample images for each algorithm respectively and build a characterization of each pair of sample images using above measure descriptors. By computing variance of image, image entropy, edge density and geometric invariants directly from sample images, we can code description feature vectors which will be fed to the learning engine of SVM. Each feature vector has 14 components totally and normalized before training.

5.2 Training

By selecting a subset of samples as a training set and the complementary subset as the test set we can build an automatic SVM classification system, which can label pair of images according to its likelihood of belonging to right matching class or not. That is to say, we construct decision function using input of feature vectors and output of class labels from samples.

The radial basis function (RBF) kernel in Eq. (14) will be used in classification, which outperforms the linear and polynomial kernel. And whole training set is trained by two best parameters which are optimized after cross validation and an exhaustive grid search. After training, a decision function of classification model is built. Thus we can test the training result using complementary test set.

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|)^2, \gamma > 0 \qquad (14)$$

5.3 Performance evaluation by classification

Analysis for performance evaluation is performed at the last step. With the input of description feature vectors based on images information, we can obtain the output label (right matching class and wrong matching class) by SVM classification and predict whether the algorithm can make good matching performance or not. From output of classification, we can judge how well a particular matching algorithm perform with respect to a certain scene and, by comparison, how well they perform with respect to one another.

6. EXPERIMENTAL RESULTS

A sequence of experiments was performed to verify our method described in this paper. Firstly, we selected SPOT 5 panchromatic images with 2.5 m resolution, ETM+ band 8 panchromatic images with 15 m resolution and SAR images with 10 m resolution as original test data. In order to undertake scene matching, all types of above images were resampled to 10 m resolution. Then we cut 100 patch samples (240×240 pixels size) from SPOT images. The patch samples, which regarded as reference images, represented typical scenes (bridge, lake, river, building, road, etc). In addition, we also cut 100 patch samples (80×80 pixels size) from ETM and SAR images respectively which located in the range of corresponding reference images and regarded as sensed images. Thus each SPOT reference image and corresponding ETM sensed image composed a pair of sample images. And each SPOT reference image and corresponding SAR sensed image composed another pair of sample images (Fig. 1). As a result, we collected 100 pairs of SPOT-ETM images and 100 pairs of SPOT-SAR images totally.

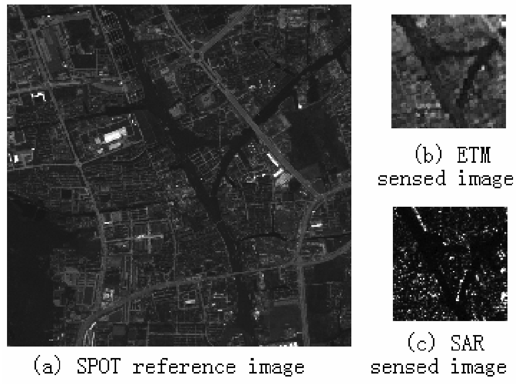

Figure 1. Pairs of sample images

We computed similarity measure descriptors from pairs of samples and coded feature vectors as input of SVM system. At the same time, we obtained matching result using NCC algorithm and CEE algorithm, which labelled as output of training data set. We use 60% of samples in training and complementary 40% in testing. That is to say, 40 samples in each group can be tested and evaluated after SVM classification. Here, confusion matrix was used to perform the analysis of overall system performances. The results of SPOT-ETM pair mode using NCC algorithm are shown in Table 1, where each column corresponds to the reference class, each row corresponds to output class decided by the SVM classification and each cell in the table gives the number of right matching class (RM Class) and wrong matching class (WM Class). We see that user's total accuracy is detected at 87.5%. Similarly, we can gain 82.5% total accuracy in SPOT-SAR pair mode using NCC algorithm (Tabel 2), 85% total accuracy in SPOT-ETM pair mode using CEE algorithm (Tabel 3) and 80% total accuracy in SPOT-SAR pair mode using CEE algorithm (Tabel 4) respectively.

NCC algorithm		Result of samples	
		RM class	WM class
Result of SVM classification	RM class	29	2
	WM class	3	6

Table 1. Confusion matrix analysis of SPOT-ETM pair mode using NCC algorithm

NCC algorithm		Result of samples	
		RM class	WM class
Result of SVM classification	RM class	17	4
	WM class	3	16

Table 2. Confusion matrix analysis of SPOT-SAR pair mode using NCC algorithm

CEE algorithm		Result of samples	
		RM class	WM class
Result of SVM classification	RM class	29	1
	WM class	5	5

Table 3. Confusion matrix analysis of SPOT-ETM pair mode using CEE algorithm

CEE algorithm		Result of samples	
		RM class	WM class
Result of SVM classification	RM class	24	4
	WM class	4	8

Table 4. Confusion matrix analysis of SPOT-SAR pair mode using CEE algorithm

In order to validate the correctness of SVM classification system output, another experiment was performed. 50 pairs of SPOT-ETM patches and 50 pairs of SPOT-SAR patches were selected randomly and fed into the trained SVM system. The results of classification are shown in Table 5 and Table 6.

NCC algorithm	Result of classification	
	RM class	WM class
SPOT-ETM pair mode	45	5
SPOT-SAR pair mode	28	22

Table 5. Performance of NCC algorithm

CEE algorithm	Result of classification	
	RM class	WM class
SPOT-ETM pair mode	43	7
SPOT-SAR pair mode	39	11

Table 6. Performance of CEE algorithm

The results of the experiment allow us to draw the following general conclusion:

(1) When the reference image and sensed image have similar intensity and texture information, NCC and CEE algorithms both work and seem to have equal matching probability.

(2) When the intensity is very different, the NCC algorithm gives bad results. On the contrary, CEE algorithm gains good results.

7. CONCLUSION

In the paper, we present a novelty objective performance evaluation approach for scene matching, which automatically trains and tests data via SVM. This approach has no a priori knowledge, and only a set of train examples for the learning step is needed, which is very important for choosing the scene matching algorithms and improving the performance of algorithms.

Another aspect of this research which should be improved is the set of measure descriptors (features) used for the SVM classification. We should select more suitable measure descriptors which can optimize computational efficiency and gain the better classification accuracy. Finally, further tests with additional data bases would be interesting in order to validate the applicability of the method to other types of algorithms and sensors.

REFERENCES

Barnea, D. I., Silverman, H. F.,1972. A class of algorithm for fast digital registration. *IEEE Trans. Comput.*, 21, pp. 179-186.

Brown, L. G.,1992. A survey of image registration techniques. *ACM Computing Surveys*, 24(4), pp. 326-376.

Canny, J. F.,1986. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8, pp. 679-698.

Coutre, S. C., Evens, M. W., Armato, S. G.,2000. Performance evaluation of image registration. *Proceedings of the 22nd Annual International Conference of the IEEE*, pp. 3140-3143.

Flusser, J.,2000. On the independence of rotation moment invariants. *Pattern Recognition*, 33 (9), pp. 1405-1410.

Hu, M.K.,1962. Visual pattern recognition by moment invariants. *IEEE Transactions on Information Theory*, 8 (2), pp. 179-187.

Moigne, J. L., Xia, W., Chalermwat, P., et al.,1998. First evaluation of automatic image registration methods. *Geoscience and Remote Sensing Symposium Proceedings*, pp. 315-317.

Sjahputera, O.,2004. *Object Registration in Scene Matching based on Spatial Relationships*. ProQuest Information and Learning Company, USA.

Suykens, J. A. K., and Vandewalle, J.,2000. Recurrent least squares support vector machines. *IEEE Transactions on Circuits and Systems-I*, 47(7), pp. 1109-1114.

Terefa, D. A., and Harada, K.,2001. A complement to the 8-point algorithm in matching uncalibrated image sequences of a scene. *Machine Graphics and Vision*, 10(1), pp. 15-27.

Vapnik, V.,1998. *Statistical learning theory*. John Wiley, New York.

Wong, R. Y.,1980. Intensity signal processing of images for optical to radar scene matching. *IEEE Trans. On Acoustics, Speech, and Signal Processing*, 28(2), pp. 260-263.

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