Cross Correlation versus Mutual Information for Image Mosaicing

Sherin Ghannam and A. Lynn Abbott

Bradley Department of Electrical and Computer Engineering

Virginia Tech

Blacksburg, Virginia, USA

Abstract—**This paper reviews the concept of image mosaicking and presents a comparison between two of the most common image mosaicing techniques. The first technique is based on normalized cross correlation (NCC) for registering overlapping 2D images of a 3D scene. The second is based on mutual information (MI). The experimental results demonstrate that the two techniques have a similar performance in most cases but there are some interesting differences. The choice of a distinctive template is critical when working with NCC. On the other hand, when using MI, the registration procedure was able to provide acceptable performance even without distinctive templates. But generally the performance when using MI with large rotation angles was not accurate as with NCC.**

Keywords—mosaicing; normalized cross correlation; mutual information

I. INTRODUCTION

Mosaicing refers to the process of combining multiple photographic images with overlapping fields to produce a single image of the whole scene. Mosaicing, also known as panographic photography or image stitching, has an extensive research literature [1-3] and several commercial applications [4-6].The automatic construction of large and high-resolution image mosaics is an active area of research in the fields of photogrammetry, computer vision, image processing, medical image, robot vision and computer graphics [7,8]. The most traditional application is the construction of large aerial and satellite photographs from collections of images.

The report is organized as follows. An overview of image mosaicing is provided in section II. Details about image registration are presented in section III. Section IV demonstrates the matching process. The methodology is presented in section V. A brief discussion about optimization is presented in section VI. Quality assessment is illustrated in section VII. Section VIII introduces the experimental results. Finally, section IX concludes the work.

II. IMAGEMOSAICING

Mosaicing, combines overlapping images to produce one composite image for a scene [1], as shown in Fig.1. The first part of an image mosaicing operation consists of identifying correspondences between some features present in both images, in order to determine the geometric transformation necessary to align the two images. This alignment operation is called *image registration*. After alignment, a composite image is created by merging or averaging pixel values of the overlapping portions and retaining pixels where no overlap occurs.
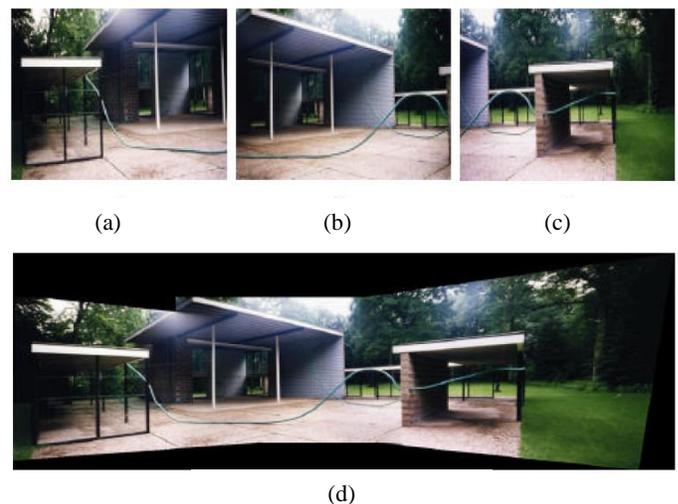

(a)   (b)   (c)

(d)

Fig. 1. Illustration of mosaicing (from [9]). (a-c) Input images. (d) Resulting mosaic.

III. IMAGE REGISTRATION

Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. The registration process requires computational methods for determining point-by-point correspondences between two images of a scene. Registration may be used to fuse complementary information in the images or to estimate the geometric and/or intensity difference between the images [10].

From corresponding positions in two images, a transformation function can be determined to get correspondences between the remaining points in the images. The aim of registration is find the transformation parameters that maximize a similarity metric or minimize dissimilarity metric between the images to be registered. The optimization problem can be formulated as follows:

$$\hat{T} = \arg \max_{T \in \tau} S(V, U, T) \qquad (1)$$

where U and V are the first and second images to be registered, T is the transformation, τ is the search space, S is

the similarity measure and $\hat{T}$ is the optimal solution. These concepts are illustrated in Fig. 2.
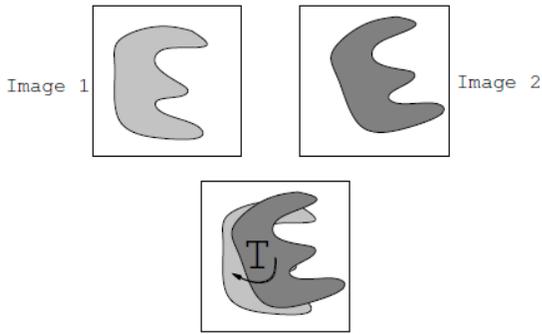

Fig. 2. Registration process (from [11]).

The success of the registration often requires that the search space τ is relatively small with respect to the input data. Large τ increases the probability of getting trapped in a local minimum.

Transformations can be rigid, depending only on translation and rotation. In that case a transformation can be represented as

$$T(X) = R X + t \qquad (2)$$

where R is a 2×2 rotation matrix with one degree of freedom, and t is a 2×1 translation vector. More generally, R can incorporate additional degrees of freedom, and the resulting affine transformation represents a composition of rotation, dilation, and shear. For either case, the transformation can be represented in homogeneous form using a single matrix M as

$$T(X) = M X \qquad (3)$$

where the single matrix M is 3×3 and X is 3×1.

The majority of registration methods consist of the following four steps: feature detection, feature matching, transform model estimation, and image resampling and transformation [3].

Feature detection refers to the detection of salient and distinctive locations in the images, such as intensity edges, corners, line intersections, etc. These features are called control points (CPs) in the literature. In the matching step, the correspondences between the features detected in the input image and those detected in the reference image are established. A detailed discussion for this crucial stage is presented in the next section.

The next step is to estimate transform model parameters, as needed in (3), using the detected correspondences. The final step is to perform that transformation, using an appropriate resampling technique (such as nearest neighbor, linear or cubic interpolation) to represent the transformed image.

IV. (DIS)SIMILARITY METRICS FOR MATCHING

Physically corresponding features can be quite dissimilar in appearance due to imaging conditions.

To identify correspondences, two major categories of matching methods are often used: area based methods and feature based methods.

A. Area (Intensity) Based Methods

Area based measures rely on computations between "windows" of pixel values in the two images. Two such methods are normalized cross correlation and mutual information. As described here, these methods provide measures of image similarity, because larger values result for corresponding points. Area based examples of dissimilarity include sum of squared difference, and sum of absolute difference [12].

An advantage of these techniques is that they can be applied to image data directly, and do not require higher-level structural analysis. But they have the disadvantage of sensitivity to intensity changes, introduced for instance by noise, varying illumination, and/or by using different sensor types.

An approach to enhance dealing with intensity changes is to use the zero-mean normalized cross correlation (NCC); also called cross covariance. It is defined as

$$NCC(i,j) = \frac{\sum_{x,y}(V(x,y) - \bar{V})(U(x-i,y-j) - \bar{U})}{\sqrt{\sum_{x,y}(V(x,y) - \bar{V})^2}\sqrt{\sum_{x,y}(U(x-i,y-j) - \bar{U})^2}} \qquad (4)$$

Where x and y are the pixel coordinates while i and j refer to the shift at which the NCC coefficient is calculated. The resulting matrix NCC contains correlation coefficients with values between -1.0 and 1.0. Note that U refers to the input image after being transformed by (3).

Mutual information (MI) is another popular matching metric used for image registration [13,14]. It is based on information-theoretic concepts, and can be considered a measure of the statistical dependency between the data sets. This metric requires the computation of joint histograms as shown in Fig.3. In the figure, n_{ij} is the number of pixels with color i in I and with color j in J. The values n_i and n_j are the marginal values, i.e., histograms of I and J.
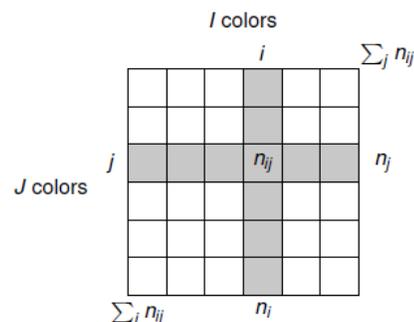

Fig. 3. Joint histogram for images I and J.

To define the MI between two images, we regard them as random variables X and Y and their intensity values at certain coordinates in the images as the joint outcome of a random experiment. MI is defined in terms of entropy as follows:

$$MI(X,Y) = H_e(Y) - H_e(Y|X)$$
$$= H_e(X) + H_e(Y) - H_e(X,Y) \quad (5)$$

where $H_e(X) = -\sum_X p_d(X)\log(p_d(X))$ represents the entropy of random variable X and $p_d(X)$ is the probability density, as estimated by a histogram. Then $H_e(X,Y) = -\sum_X \sum_Y p_d(X,Y)\log(p_d(X,Y))$ represents the joint entropy between the two random variables X and Y.

In information theory, entropy is considered to be a measure of the uncertainty in a variable. For illustration, example probability distributions and their associated entropy values are shown in Fig. 4. Flatter distributions represent higher levels of uncertainty in the output of an experiment.

Typically, the joint probability distribution of two images is estimated as a normalized joint histogram of the intensity values. The marginal distributions are obtained by summing over the rows or over the columns of the joint histogram. The normalized joint histogram is calculated by counting the number of pixels having an intensity value i in the first image and an intensity value j in the second image at the same location and denoting it by n_{ij}. At each location in the first image, the pixel intensity value i will be examined and its corresponding value j in the same position in the second image. This will increase the counter n_{ij} by one. This process is repeated for each pixel. To obtain the normalized joint histogram, all values of n_{ij} should by divided by the sum of these values.
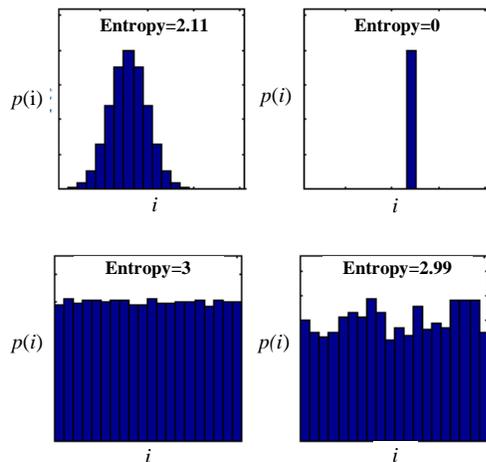

Fig. 4. Examples of probability distributions with associated entropy values.

Registration is achieved by adjustment of the relative position and orientation until the MI between the images is maximized. Some studies have shown the use of MI can give good performance even for multimodal situations [13].In such cases, intensity values are not linearly related so the cross correlation is not likely to succeed.

B. Feature Based Methods

These methods are preferred when the local structural information is more significant than the information carried by the image intensitiesal one. Features should be distinctive, distributed well over the images, and efficiently detectable in both images. These features can be based on regions, lines, and points.

Feature detection from regions relies on the ability to extract a useful subset of an image, possibly by applying a thresholding operation to a high-contrast image. Examples of region features are centroid, area, and elongation.

Line features can result from edge detection followed by line fitting. Common edge detection methods, such as the Canny technique or the Laplacian of Gaussian, result in a set of points in the image. A line-fitting algorithm such as the Hough transform can result in a set of lines, and their properties can be used as features for matching.

Point features often rely on the detection of intensity edges or corners in an image. Fig. 5 shows corner correspondences that have been detected in two images [9].
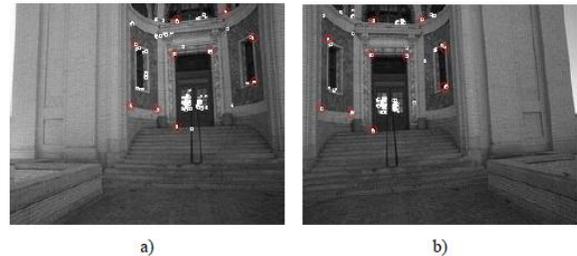

Fig. 5. Corner correspondences between two images (from [9]).

V. METHODOLOGY

This section illustrates the problem of creating an image mosaic from two overlapping 2D images of the same 3D scene. The NCC and MI similarity measures are used for registration, and the results are compared.The steps are summarized in Fig. 6.

A. Image Preprocessing

If image involves more than one color band, e.g. RGB, typically only one band is taken into consideration while performing the matching process. But after finding the optimal transformation parameters, all color bands will be processed during the transformation (alignment) stage to produce a color mosaic.

B. Matching

Given the two images, the task is to find correspondences between them. This is done by performing registration using two area-based similarity measures, NCC and MI. The position at which either NCC or MI is maximized will be stored.

To accommodate rotation, this process will be repeated at several rotation angles of the original image. This search

interval can be reduced using an optimization technique which will be illustrated in the next section.

Applying this matching process over the whole image consumes a lot of time especially for large images. This process can be accelerated using small template(s) taken from each of the images. The matching process will be carried out using these templates. The template-matching process that finds the correspondence produces more reliable matches if the selected templates are locally unique [15]. A template that is relatively homogeneous may easily lead to false matches.

Fig. 6.   Flow chart of image mosaicing system.

C. Transformation

After using image templates to determine the geometric transformation T that is needed, the entire input image is transformed using (3). This process is sometimes called alignment.

D. Interpolation

Typically, a transformation T will require the computation of new pixel values using several pixels from the original image. This step can be performed by averaging pixel values locally, although other techniques such as spline-fitting have also been employed.

VI.   OPTIMIZATION

Optimization techniques often rely on the maximization or minimization of an objective (cost) function[16]. Many optimization techniques can suffer from finding solutions that correspond to a local optimum of the objective function, instead of the solution that is the best overall.

For optimizating functions of n variables, many algorithms work by doing a sequence of 1D optimizations. For the case of 1D minimization, as illustrated in Fig. 7, it is possible to subdivide a given range [a,c] iteratively in an attempt to find the optimum solution.

Pseudocode for the process is given in Fig. 8. Line search techniques proposed in the literature differ in the way of calculating the value of the intermediate point b. For example, in the golden search method, b is either $\left(\frac{1}{2} - \frac{\sqrt{5}}{2}\right)$ or $\left(\frac{\sqrt{5}}{2} - \frac{1}{2}\right)$ times the distance between a and b.
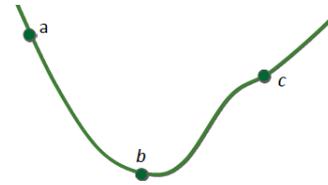

Fig. 7.   Subdividing process for 1D optimization.

% Line search algorithm stops when reaching a prespecified tolerance (TOL) or maximum number of iterations K

$$while\big((c-a) > TOL \mid iterations < K\big)$$
$$if\ f(b) < f(a)\ \&\ f(b) < f(c)\ \ then$$
$$if\ f(a) > f(c)\ then$$
$$a = \frac{a+b}{2}$$
$$else$$
$$c = \frac{b+c}{2}$$
$$endif$$
$$else$$
$$if\ f(a) < f(b)\ \ then$$
$$dx = b - a$$
$$c = b$$
$$b = a$$
$$a = b - dx$$
$$else$$
$$dx = c - b$$
$$a\ = b$$
$$b = c$$
$$c = b + dx$$
$$endif$$
$$endif$$
$$endwhile$$

Fig. 8.   Pseudocode for basic line search.

In the problem of image registration, the only method that guarantees a globally optimum solution is an exhaustive In the

problem of image registration, the only method that guarantees search over the entire image. Although computationally demanding, such a global search is often used if only translations are to be estimated. In case of transformations with more degrees of freedom or in case of more complex similarity measures, more sophisticated optimization algorithms are required [12].

VII. QUALITY ASSESSMENT

Different metrics can be used to measure the quality of the mosaicing result [17]. Mean square error (MSE) and peak signal-to-noise ratio (PSNR) are commonly used. They are applied on the overlapping region O:

$$MSE = \frac{\sum_{(x,y)\in O}\big(I_2(x,y) - I_1(x,y)\big)^2}{N} \quad (6)$$

$$PSNR = 10\, log_{10} \frac{(255)^2}{\sum_{(x,y)\in O}\big(I_2(x,y) - I_1(x,y)\big)^2} \quad (7)$$

A reconstruction error (E) measuring the mean of the absolute intensity differences between two successive images on the overlapping area (O) has also been used [18]. It is defined as

$$E = \frac{1}{|O|} \sum_{(x,y)\in O} |I_2(x,y) - I_1(x,y)| \quad (8)$$

Exposing one image to changes in luminance or noise greatly affects these parameters. This means that low PSNR or high MSE or high E does not always mean poor registration.

CVLab [19] has suggested an evaluation methodology for the comparison of image mosaicing algorithms. The idea is to compare mosaics to their ground truth versions. This work was inspired by the work of in the performance evaluation of stereo reconstruction algorithms [20].

VIII. EXPERIMENTAL RESULTS

Mosaicing techniques using NCC-based and MI-based similarity measures were implemented using Java, and were applied to two different image pairs to investigate their strengths and shortcomings. The first image pair, shown in Fig. 9(a-b), is from a football field. The images contain some obvious distinctive parts against a background that is largely homogenous. The second image pair is from remote sensing. For these experiments, only rigid transformations, translation and rotation, were considered.

The first two images overlap by nearly 8%. Templates for matching were obtained by applying a thresholding operation, and they are shown in Fig. 10. Because the two templates are fairly small in size, searching for the optimal translations and is relatively fast. But repeating this process for many possible rotation angles is computationally intensive, especially for the MI-based technique. In order to reduce the computation time, a line search (as described previously) was applied to the rotation-angle search space.

The second image pair is shown in Fig. 11(a-b).The two images overlap by nearly 25%.Templates for matching were obtained in this case, as shown in Fig. 12, through a cross correlation using binary templates that were extracted from the two images.
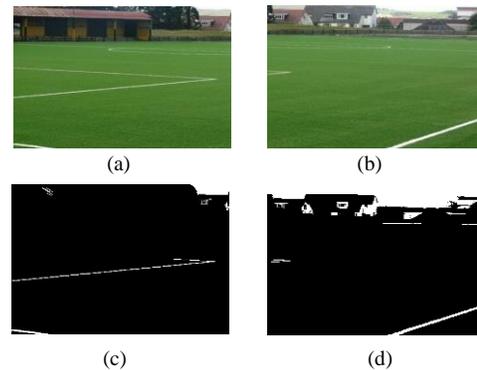

Fig. 9. Thresholding operation on first image pair. (a-b) First and second images. (c-d) Binary images.
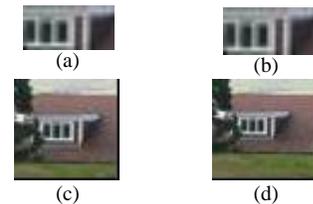

Fig. 10. Template selection. (a-b) Selected objects from binary images. (c-d) Corresponding templates.
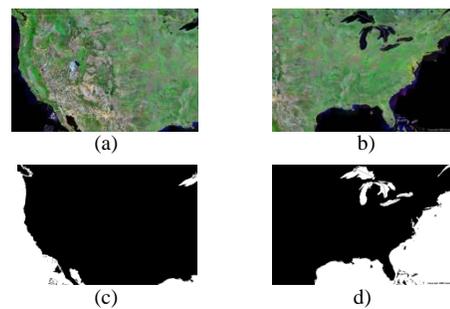

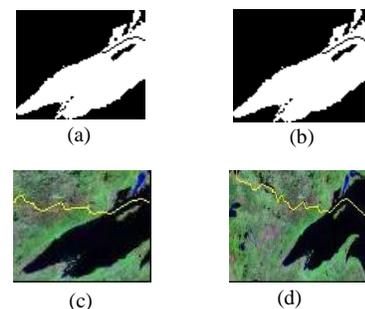Fig. 11. Thresholding operation on second image pair. (a-b) First and second images. (c-d) Binary images.

Fig. 12. Template selection. (a-b) Selected objects from the two thresholdedimages. (c-d) Corresponding templates.

Mosaicing results for different rotation angles are shown in Table I. The two input images are displayed in the first column, the two chosen templates are in the second column and the third column contains the resulting mosaics. Values of NCC, MI, and PSNR for the overlapping areas are displayed under each result. A number of experiments were carried out to test the performance of the mosaicing process under non-modeled variations such as intensity manipulation, shearing, and Gaussian noise addition. The resulting mosaics after applying these operations for zero rotation angles are shown in Table II.

To clearly compare the performance of the two techniques, graphs showing the computed rotation angles from each technique at each case versus the actual rotation angles are plotted in Figs.14 to 21.

From Fig. 13and Fig. 17, it can be seen that registration using MI may have a considerable amount of error at large rotation angles, while the performance of the NCC-based technique was better. If there is a substantial difference in luminance between the two images, these errors increase. For example, the MI-based technique successfully deduced the correct angle when the actual rotation angle was -30 in the second test set before increasing the brightness of one image. After changing the luminance, the MI-based test yielded an angle of 71°, as shown in Fig. 18. On the other hand, NCC succeeded in finding angles very close to the correct ones.

Adding Gaussian noise (with a variance of 12) also affected the resulting mosaics and produced some errors as shown in Fig.19. The errors in the case of NCC-based were around 5° while for the case of MI, errors were fairly large at some large rotation angles. The same scenario happened after applying shearing with factors of 0.15 and 0.35 as shown in Fig.20. Shearing caused almost all computed angles to be incorrect by about 5°, plus some dramatic errors in the case of MI-based at large rotation angles.

Althought it seems that the NCC-based technique is superior to the MI-based technique, especially at large rotation angles, there are some situations that the NCC-based technique could not handle and MI-based performance was much better.

One of these situations corresponds to the case that the chosen templates do not have distinctive details (i.e. nearly homogenous texture). Table III shows some results of the mosaicing techniques based on templates that are not distinctive. Even at small rotation angles such as 10°, the NCC-based technique could not extract accurate registration parameters, as shown in Fig. 21. However, the MI-based technique succeeded.

TABLE I.   SOME RESULTS FOR DIFFERENT ROTATION ANGLES.

Test Image 1: Horizontal and vertical translation and zero rotation

NCC-based result
NCC=0.997, MI=1.84, PSNR=29.45

MI-based result
NCC=0.997, MI=1.84, PSNR=29.45

Test Image 1: Horizontal and vertical translation and 10° rotation

NCC-based result
NCC=0.996, MI=1.85, PSNR=29.06

MI-based result
NCC=0.996, MI=1.85, PSNR=29.06

Test Image 1: Horizontal and vertical translation and 50° rotation

NCC-based result
NCC=0.996, MI=1.85, PSNR=29.1

MI-based result
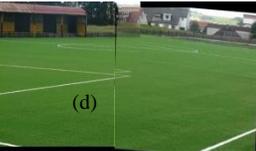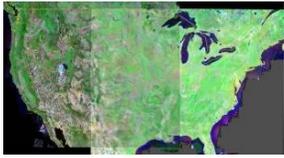NCC=0.996, MI=1.185, PSNR=29.1

Table I, continued

TABLE II.   RESULTS OF SOME EXPERIMENTS.

Test Images 2: Horizontal and vertical translation and zero rotation

NCC-based result
NCC=0.999, MI=2.66, PSNR=35.46

MI-based result
NCC=0.999, MI=2.66, PSNR=35.46

Test Images 2: Horizontal and vertical translation and 10° rotation

NCC-based result
NCC=0.999, MI=2.69, PSNR=35.22

MI-based result
NCC=0.993, MI=2.35, PSNR=27.69

Test Images 2: Horizontal and vertical translation and 50° rotation

NCC-based result
NCC=0.996, MI=2.33, PSNR=30.13

(d)

MI-based result
NCC=0.967, MI=1.52, PSNR=20.91

Intensity Adjustment

NCC-based result | **MI-based result**

NCC-based result | **MI-based result**

Noise Addition

NCC-based result | **MI-based result**

NCC-based result | **MI-based result**

Shearing

NCC-based result | **MI-based result**
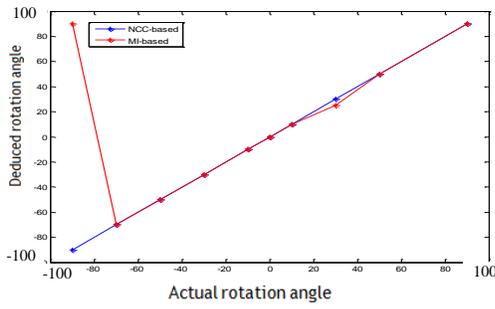

NCC-based result | **MI-based result**

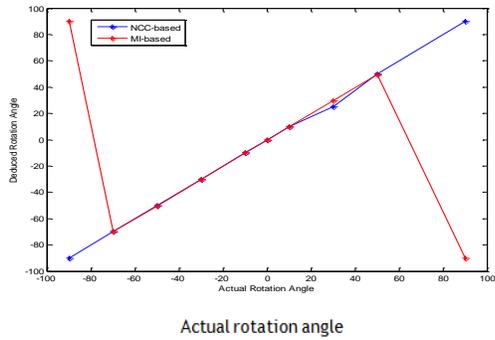Fig. 13. Computed rotation angles versus actual rotation angles for test image 1.

Fig. 14. Computed rotation angles versus actual rotation angles for test image 1 after intensity adjustment.
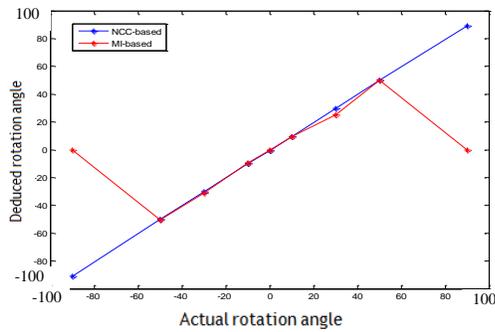

Fig. 15. Computed rotation angles versus actual rotation angles test image 1 after noise addition.
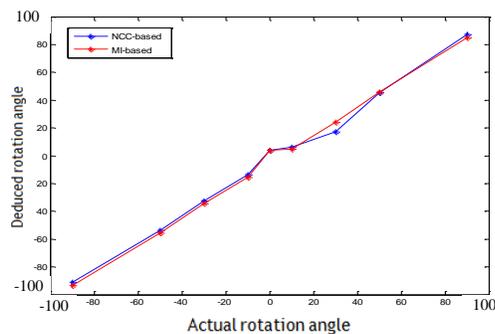

Fig. 16. Computed rotation angles versus actual rotation angles for test image 1 after shearing.
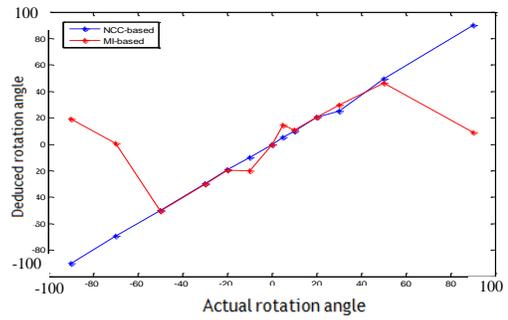

Fig. 17. Computed rotation angles versus actual rotation angles for test image 2.
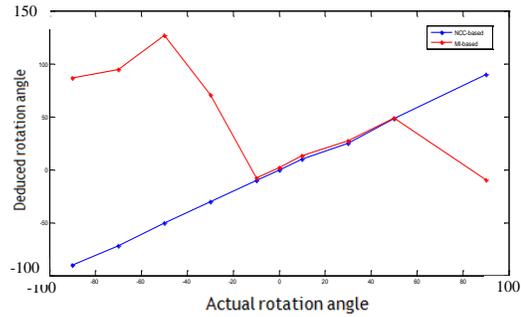

Fig. 18. Computed rotation angles versus actual rotation angles for test image 2 after intensity adjustment.
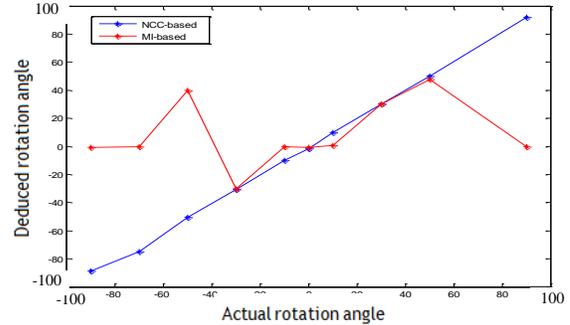

Fig. 19. Computed rotation angles versus actual rotation angles for test image 2 after noise addition.
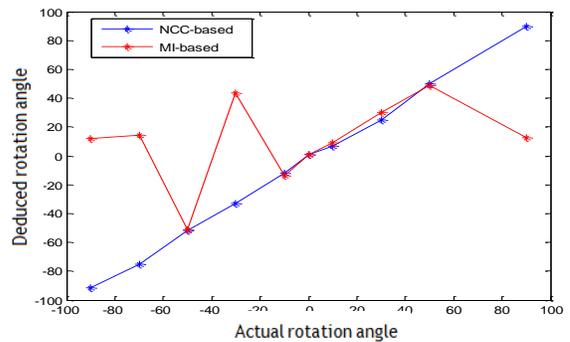

Fig. 20. Computed rotation angles versus actual rotation angles for test image 2 after shearing.

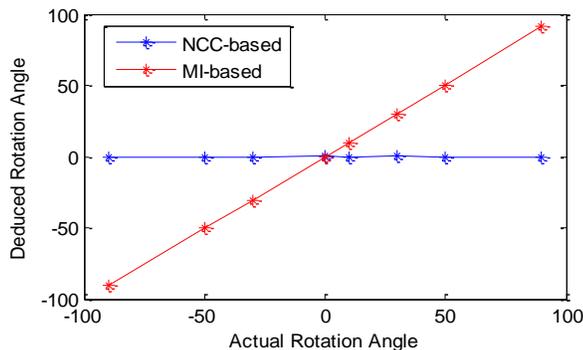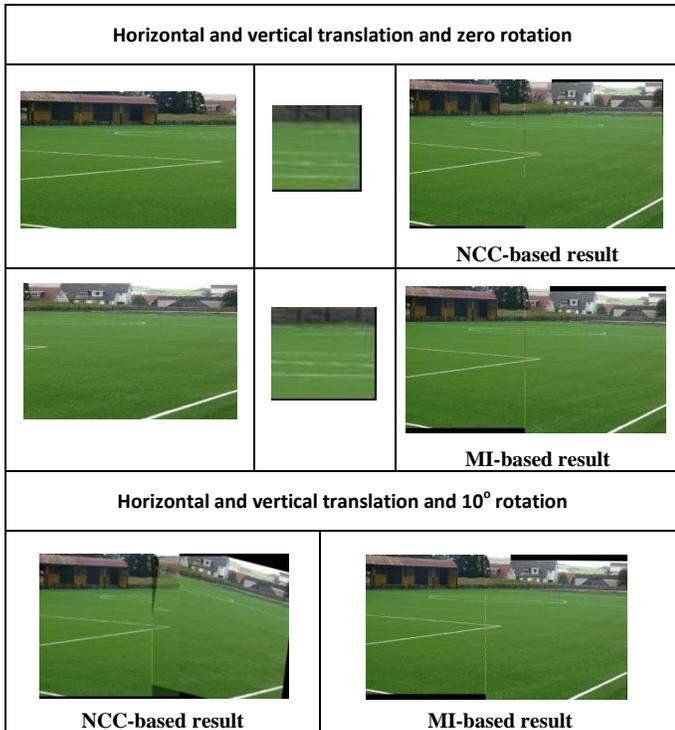TABLE III.   RESULTS BASED ON INDISTINCTIVE TEMPLATES.

Horizontal and vertical translation and zero rotation
  NCC-based result
  MI-based result
Horizontal and vertical translation and 10° rotation
  NCC-based result   **MI-based result**

Fig. 21. Computed rotation angles versus actual rotation angles based on indistinctive templates.

IX.   CONCLUSION

This paper has provided a comparative evaluation of the performance of template-based mosaicing techniques using two common similarity measures:normalized cross correlation (NCC) and mutual information (MI). Their performance has been tested at different rotation angles and under some un-modeled distortions and intensity variations.

The results indicate that NCC-based and MI-based mosaicing techniques have very close performance in many cases, but NCC-based performance is usually better for large rotation angles.

To provide faster results, small registration parameters were obtained using small image templates. When working with NCC-based mosaicing, steps should be taken to ensure that these templates are distinctive. Otherwise, the system may fail to provide reliable rotation angles. In the results shown here, the MI measure was less sensitive to the choice of image templates.

REFERENCES

[1] R. Szeliski, "Image alignment and stitching: A tutorial", Technical Report MSR-TR-2004-92, Microsoft Research, 2004.

[2] D. Milgram, "Computer methods for creating photomosaics",*IEEE Transactions on Computers*, C-24 (11):1113–1119, 1975.

[3] M. Brown and D. Lowe, "Recognising panoramas", Proc.9th Intl. Conference on Computer Vision, vol. 2, pp. 1218–1225, 2003.

[4] S. E. Chen, "Quick Time VR—An image-based approach to virtual environment navigation", Proc. SIGGRAPH'95, pp. 29–38, 1995.

[5] Realviz, http://www.realviz.com

[6] http://www.microsoft.com/products/imaging

[7] R. B.Inampudi, "Image mosaicing", Proc. IEEE Int.Geoscience and Remote Sensing Symposium(*IGARSS'98*), vol. 5, pp.2363 – 2365, 1998.

[8] H. Shum, and H. and R. Szeliski, "Construction and refinement of panoramic mosaics with global and local alignment", Proc. IEEE Intl. Conf. on Computer Vision, pp. 953-958, 1998.

[9] A. Rocha, R. Ferreira and A. Campilho, "Image mosaicing using corner detection", Proc. Ibero-American Symposium on Pattern Recognition, 2000.

[10] L. G. Brown, "Survey of imageregistration techniques", *ACM Computing Surveys*, vol. 24, no. 4, pp. 325-376, 1992.

[11] http://www.creatis.insa-lyon.fr/~srit/tete/2012_master_eeap_si_m5.pdf

[12] B.Zitováand J. Flusser, "Image registration methods: a survey*"*, *Image and Vision Computing Journal*, vol. 21, no. 11, pp. 977-1000, 2003.

[13] P. Viola, "Alignment by maximization of mutual information", PhD thesis, MIT, 1995.

[14] A. Collignon, "Multi-modality medical image registration by maximization of mutual information", Ph.D. Thesis, Catholic University of Leuven, Leuven, Belgium, 1998.

[15] A. Ardeshir Goshtasby, *2-D and 3-D Image Registration: For Medical, Remote Sensing, and Industrial Applications*, John Wiley & Sons, 2005.

[16] D. A. Wismer and R. Chattergy, *Introduction to Nonlinear Optimization*, North Holland, 1978.

[17] A. Bevilacqua, A. Gherardi, and F. Piccinini, "Quantitative quality assessment of microscopic image mosaicing", *Journal of World Academy of Science, Engineering and Technology (WASET)*, vol. 47, no. 48, 283-286, 2010.

[18] E. Zagrouba, W. Barhoumi, and S. Amri, "An efficient image-mosaicing method based on multifeature matching," *Machine Vision and Applications*, vol. 20, pp. 139–162, 2009.

[19] http://www.vision.deis.unibo.it/MosPerf/MosPerf-Home.aspx

[20] http://vision.middlebury.edu/stereo/