Reflective Particle Tag for Arms Control and Safeguards Authentication

Antonio Gonzales, Nairong Wang, Aleksandra Faust, Kurt Larson
Advanced Information Systems Organization 5534

Jack Bartberger, Karl Horak, Nathan Schanfein, Manuel Trujillo, Keith Tolk, and Peter Merkle
Nuclear Materials Monitoring and Advanced Technologies Organization 6722

Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Introduction

The Reflective Particle Tag (RPT) is a next generation of tag technology developed with the goal of providing both a unique identifier and visual evidence of tampering. They are resin tags infused with reflective crystals that are placed over sensitive seals. Our goal is to develop a verification routine based on image processing that verifies the original set of images (gallery), of an RPT, against a set taken at a later time (probe). Each set contains images taken by illuminating the RPT from a number of orientations, creating different views due to crystal reflections. In this paper, we describe an algorithm for RPT verification using state-of-the-art computer vision techniques and discuss future work to help identify low-evidence tampering.

RPT Verification Algorithm

We propose a multi-step approach for RPT image verification. We start by finding the corresponding image pairs from the gallery and probe image sets that share the same lighting orientation. For each corresponding pair, we attempt to find a geometric relationship between the images using techniques from the field of wide baseline matching (WBM). If a geometric relationship is found, it is then tested to ensure that it meets a rigorous criterion. If the geometric relationship meets this criterion, the two images are subtracted to look for any evidence of tampering. Finally, we ensure that all image pairs are verified correctly. A flowchart of our algorithm is shown in Figure 1.

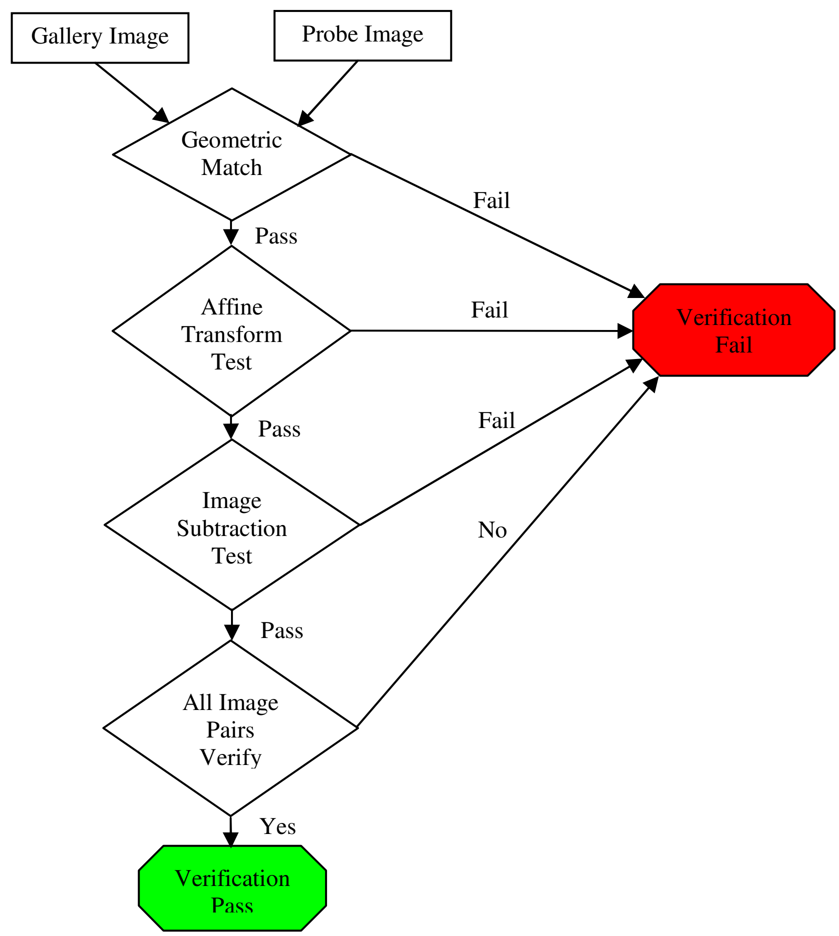

Figure 1: Flow chart for RPT image-based verification

Currently, we have only implemented and tested the geometric matching and affine transformation testing steps. In the following sections we discuss the implementation and testing of these steps. We will discuss our ideas on the remaining steps in the Future Work section.

Wide Baseline Matching

WBM is a widely studied and validated paradigm of computer vision that locates identical features across a set of images of the same object. The locations of corresponding features are then used to solve for the geometric relationship between the two images. While varying widely in design, WBM algorithms all share three key elements:

1. feature detectors that can repeatedly locate the same feature in different images under varying conditions;
2. feature descriptors that characterize features, and their local area, in a way that will return the same or very similar values under varying viewing conditions, and
3. methods that identify feature correspondences and their associated geometric organization, which ensure a high probability match.

WBM applies well to the RPT problem because it will reliably fail if the RPT is completely different or severely changed, but is effective for small, expected changes such as weathering or small, inadvertent damage. The next three sections detail the specific algorithms we used for feature detection, descriptor creation, and geometric correspondence.

Feature Extraction

We chose the maximally stable extremal region (MSER) algorithm as our feature detector. The MSER algorithm searches for connected areas of common intensity in the image[1]. These areas remain detectable across a wide array of image transformations and illumination changes. This algorithm has been shown to provide the best, general performance among the current slate of feature detection algorithms [2].

The MSER algorithm works by iterating through each pixel intensity value [0-255] and identifying all pixels with an intensity that is less than or equal. Neighboring pixels that pass this test are grouped together, and the size of the group is tracked across the iteration. When the iteration is complete, the algorithm examines the growth curve for each pixel group and identifies points in the curve when the group maintained a stable size over a large range of pixel values. For each identified point in the curve, the set of pixels that belonged to the group, at that point, are selected as a feature. Each feature is extracted by taking the region of the image inside the best-fit enclosing ellipse for the identified group of pixels. The algorithm is then repeated while iterating intensity values in the opposite direction [255-0] to ensure that both dark and light MSERS are extracted. Figure 2 displays a RPT image marked with a small subset of the extracted MSERS marked in red.

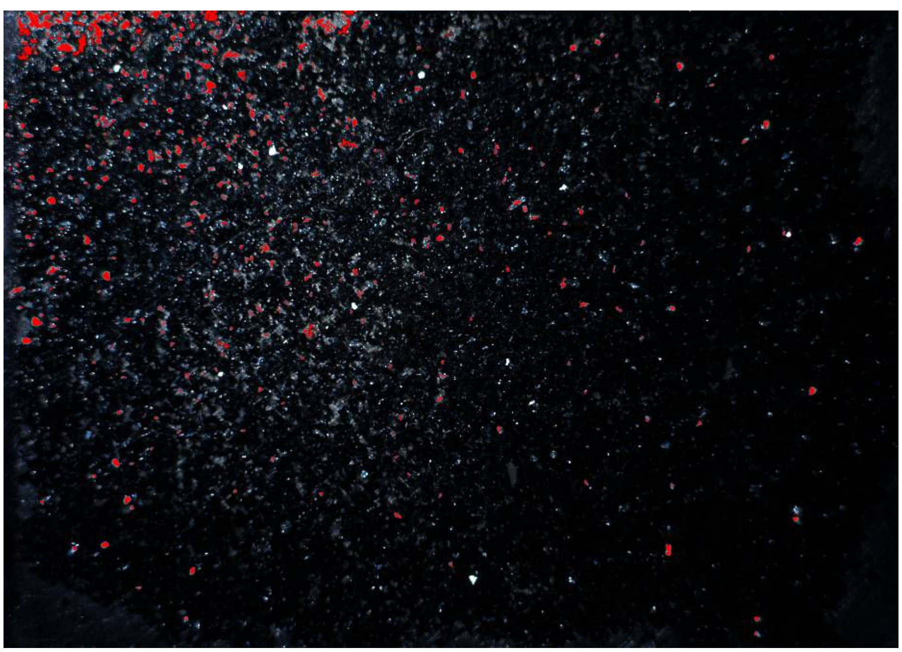

Figure 2: RPT image with detected MSERS marked in red (1,000 of 12,215 shown).

Descriptor Creation and Matching

Given sets of extracted features for two images, the next step is to find corresponding features between the images. We can not compare the features directly due to translations, rotations, and scaling that may exist between the images. Therefore, we convert the extracted features to a form that removes these issues. This is accomplished by converting the feature region into a vector of values called a descriptor. The standard method for creating descriptors is the scale invariant feature transform (SIFT) [2][3]. SIFT descriptors are 128 value vectors, independent of image rotation, scale, and illumination, and therefore, directly comparable.

To create a SIFT descriptor, we take the image feature and reshape it into a 16x16 square patch, effectively removing scale. For each pixel in this patch, we calculate the magnitudes and orientations of the gradients. A histogram is created for the orientations and the largest bin is designated as the dominant orientation. The feature patch is rotated to align with the dominant orientation which removes rotation. The transformed patch is divided into 16 individual blocks of 4x4 pixels and an 8-directional histogram is created for each by binning the gradient magnitudes and orientations. These 16 separate 8-directional histograms then combined into a single 128 value vector. The descriptor values are normalized to help reduce the effects of illumination changes.

Each image is now defined by a set of SIFT descriptors and their associated pixel position. We generate correspondences between the images by identifying the SIFT descriptors most closely matched. Descriptor comparison is done via simple n-dimensional Euclidean distance. Only descriptor matches that meet a certain distance threshold are considered good matches. In addition, we require that the best match for a descriptor must be 20% better than the second best match. This helps avoid confusion when multiple descriptors look very similar. This set of feature matches becomes the input to the final step of WBM which searches for a geometric relationship between the images.

Geometric Matching

We use the random sample consensus (RANSAC) algorithm to solve for a geometric relationship between a set of corresponding feature matches. In the case of RPT verification, the geometric relationship we are searching for is an affine transform that allows for translation, scaling, and rotation. An affine transformation can be generated for any set of three positional correspondences using a standard least-squares approach.

RANSAC starts by randomly selecting three feature matches from the set and generating an affine transformation based on the corresponding pixel positions. All remaining feature matches are checked against this affine transform by projecting each pixel position from the first image, via the transform, into the second. The location of the projected pixel position is compared to its expected location using Euclidean distance. If that distance falls under a defined threshold, in our case one pixel, it is added to the consensus set of supporting feature matches. This process is iterated to ensure high confidence that a good model will be sampled if it exists. The affine transform with the largest consensus set is returned as the solution.

RANSAC will always produce a solution with a minimum consensus set size of three. Usually, only RANSAC results with a large consensus set size are considered good. Therefore, we need to perform an additional step to verify the RANSAC output.

Affine Transformation Filtering

The final step in our RPT verification method is to ensure that the solution produced by the geometric matching algorithm meets a rigorous criterion. In RPT verification, we expect only very small transformations between images because fiducials are used to line up the camera during image acquisition. We apply a polar decomposition to break the affine transformation back into its constituent translation, scale, and rotation components. These values are compared against thresholds to ensure that they conform to our expectation of a limited amount of translation, very small rotation, and almost no scaling of any kind.

RPT Verification Experiments and Results

We were provided with a database of 4932 images made up of multiple images of 244 RPTs taken at four different lighting orientations: upper left, lower left, upper right, and lower right. An RPT image should only verify against another image if it is of the same RPT taken at the same lighting orientation.

Test 1

In our first test, we explored the performance of the geometric matching technique by itself. We chose one image for each of the 244 RPTs, at a random selected lighting orientation, and attempted verification against all images in the database. This resulted in 1,203,408 RPT verifications. We create a measure called "match percentage" by taking the ratio of the consensus set size of the geometric solution to the total number of features extracted from the probe image. This allows us to fairly compare the results of the geometric matching algorithm across different images. The summary statistics for the Test 1 are shown in Table 1.

	Number of Examples	Mean match %	Match % STD	Min Match %	Max Match %
Matching RPTs	1005	45.47%	6.4%	3.28%	55.42%
Non-Matching RPTs	1192297	0.02%	5.2e-5%	0.01%	0.46%

Table 1: Results of Verification Tests without Affine Transformation Filtering

In these results, we use the term "Matching RPTs" to refer to all verifications that should pass because the images are of the same RPT at the same lighting orientation. The term "Non-Matching RPTs" refers to all verifications that should fail. The results show that the geometric matching algorithm performs very well. The mean match percentage for matching RPTs is clearly more distinct than that of the non-matching RPTs. Unfortunately, when comparing the minimum match percentage for matching RPTs against the maximum match percentage of non-matching RPTs, we get a much smaller range of 0.46% to 3.28%. Upon investigation of the matching RPTs with low match percentages, we found that these images all had focus problems which reduced the number of feature correspondences.

We generated a receiver operator characteristics curve (ROC) by varying match percentage required for a verification to pass. We expect 1005 true positives and 1192297 true negatives. The resulting ROC curve is shown in Figure 2.

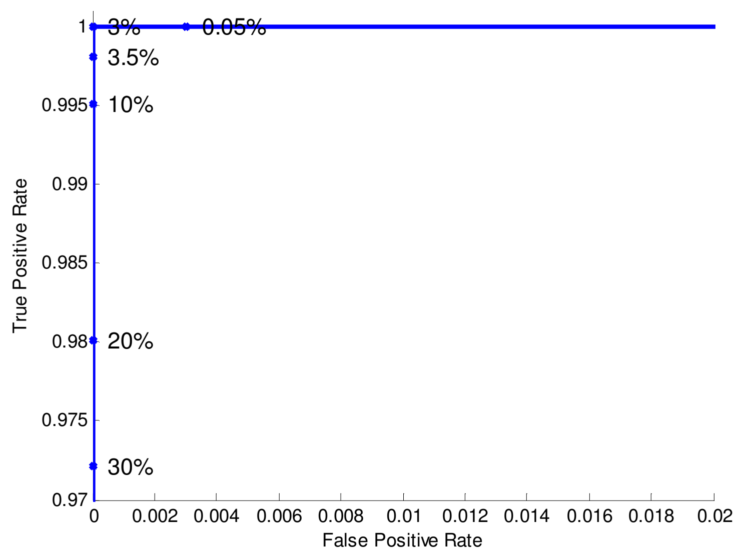

Figure 2: ROC curve for RPT Verification without Affine Transformation Filtering

As expected from the summary statistics, we obtain perfect performance for all thresholds in the range [0.46%, 3.28%]. These results are very promising but do illustrate the need for a more rigorous test to account for the possibility of getting low match percentages on true matches due to imaging issues.

Test 2

In our second test, we investigated the two different types of non-matching RPTs: verifications between RPTs that are completely different, and verifications between the same RPT with a different lighting orientations. We broke down the results from Test 1 for each of these categories. The results are shown in Table 2.

	Number of Examples	Mean match %	Match % STD	Min Match %	Max Match %
Same RPT, Different Lighting	3685	0.02%	5.04e-5%	0.011%	0.059%
Different RPTs	1188612	0.02%	5.21e-5%	0.01%	0.45%

Table 2: Same RPT, Different Lighting vs. Different RPT Matching Results

We performed a T-Test to determine if the two populations where statistically different. The result was a p-value of 0.5089 at a significance level of 5%. This means that the populations are statistically indistinguishable. This means that changing the lighting orientation by 90 degrees produces as much appearance difference as completely changing the tag.

Test 3

In our final test, we repeated the same procedure as in Test 1 and then tested the resulting affine transform using the following criteria:

- x/y translation < 100 pixels
- scale change of less than 2%
- rotation of less than 2 degrees

All 1005 matching RPT verifications passed the affine transform test. We did not have a single non-matching RPT verification pass the affine transformation test. We get perfect verification performance with any match percentage greater than 3.28%. We can not produce any false positives even when completely ignoring the match percentage. There remains a small possibility that the geometric matcher will incorrectly produce an affine transform that meets our criteria for two RPTs that should fail. Due to this fact, we can not rely on the affine transformation test alone. These results give us high confidence that the combination of our geometric matching technique with the affine transformation test produces robust RPT verifications.

Future Work

While we are confident our algorithm correctly verifies if two images are of the same RPT at the same lighting, it can not determine if there are small areas of the RPT that have been changed. To fix this weakness, we will investigate extending the verification algorithm. First, we plan to investigate aligning the images, via the solved affine transform, and subtracting the images. We will perform statistical analysis of the distribution of difference values to automatically detect areas of the image that are likely to contain non-random changes. In addition, we will investigate building templates out of this statistical information to help alert the user to the likely causes of detected change. The final step will be to combine the verification results for the images at each lighting orientation for a given RPT. By combining the results from across the image set, we will obtain even more confidence in the results.

References

[1] J. Matas, O. Chum, M. Urban, and T. Pajdla, *“Robust wide baseline stereo from maximally stable extremal regions,”* BMVC, pp. 384-393, 2002

[2] K. Mikolajczyk and C. Schmid. *“A performance evaluation of local descriptors,”* In Proc. IEEE CVPR, Volume 2, pp. 257-263, 2003.

[3] D. Lowe. *"Distinctive image features from scale-invariant keypoints,"* International Journal of Computer Vision, 60, 2, pp. 91-110, 2004.