# PLANAR-PROJECTIVE SUMMATION INVARIANT FEATURES FOR CAMERA NETWORKS

*Kerry R. Widder[1], Wei-Yang Lin[2], Nigel Boston[1], Yu Hen Hu[1]*

[1]University of Wisconsin-Madison, Dept. of Electrical and Computer Eng.,
1415 Engineering Drive, Madison, WI, 53706 USA
[2]National Chung Cheng University, Dept. of Computer Science and Information Engineering,
EA 310, 168 University Rd., Min-Hsiung, Chia-Yi, Taiwan
*widder@wisc.edu, wylin@cs.ccu.edu.tw, boston@math.wisc.edu, hu@engr.wisc.edu*

## ABSTRACT

Recently a novel family of geometrically invariant features, called summation invariant, has been developed and applied to object recognition. The range of this family of features is expanded here beyond the Euclidean and affine transformation groups to planar projective transformations. Whereas other methods require small changes in view, or collinear points, this method removes those limitations and allows recognition of general planar objects over wide ranges of viewpoint. The derivation of these new features requires the innovation of deriving the invariants in the homogeneous coordinate space, yet yields results formulated in terms of Cartesian coordinates. Simulations demonstrate the effectiveness of this new approach to object recognition under projective transformations like those encountered in camera networks.*

***Index Terms***— *summation invariant; moving frame; geometrically invariant feature; camera network; projective invariant*

## 1. INTRODUCTION

Invariant features are important tools in the pattern recognition toolbox. Objects to be recognized in images usually are not guaranteed to be in the same location, to have the same orientation, the same size, nor even to have the same shape. Thus, having a descriptive feature that is invariant to geometric transformations - like translation, rotation, scale or shear - is highly desirable. The number of transformation types included in the feature will determine how general its application is. This generality, however, has a price. As generality increases, so does the difficulty in finding invariant features.

In today's world, computer vision applications like surveillance and security have become more important. These types of applications often involve multiple cameras observing the environment from widely different perspectives. Object recognition within this context is made more difficult by the need to use projective transformations.

Many previous works have made the simplifying approximation of substituting affine transformations for projective. This approximation is reasonable if the object is located far from the camera. This assumption was made in [1], where invariant features were derived for curves under affine transformations after first transforming the curve at each point to a canonical coordinate system. These invariants required sixth order derivatives in general, or as low as second order if some feature correspondences were known. The use of higher order derivatives makes this approach sensitive to noise.

The same assumption was made in [2], where a different approach was taken to the recognition problem. Instead of deriving geometrically invariant features, each curve was normalized to a standard position that is invariant to any affine transformations of the curve. This process utilized global features, moments and Fourier descriptors, to perform the normalization.

The classic projective invariant is the cross-ratio. This invariant feature is constructed from four collinear points. This feature has been applied to planar polygons [3], to general planar curves using the curve's convex hull to find corresponding points [4], and to the retrieval of images of buildings using a cross-ratio histogram feature [5]. Another related feature is a set of five points on a plane. This invariant requires the construction of additional points, and yields two sets of four collinear points and their associated cross-ratios. These features are limited by their co linearity requirement. Another limiting factor in their usefulness is that they produce one value, which is good for invariance, but not as good for discrimination.

Integral invariants have been applied to the projective transformation group, but no analytical formula could be found for them and numerical solutions had to be used to compute them [6].

Recently, a new family of geometrically invariant features, called summation invariant, has been developed [7-10]. This family of features utilizes potentials consisting of summations as coordinates in a prolonged jet space. They are not sensitive to noise, can be generated systematically and can be implemented semi-locally, providing multiple values for better discrimination. These invariants are derived over a specific transformation group (e.g., Euclidean, affine, projective) and a specific dimension (e.g., 2D, 3D), and are invariant to transformations within that group. Previous works have shown summation invariants for Euclidean and affine transformations in both 2D and 3D with applications to shape recognition and face recognition.

* This research is partially supported by the United States National Science Foundation under grant number CCF-0434355, and partially supported by the National Science Council, Tiawan (Grant No. 96-2221-E-194-055).

This paper initiates the extension of summation invariants to the projective transformation group, providing a more general invariant than the cross-ratio. The previous application to face recognition was a situation where the object itself is transformed. Here, the object is fixed, but the observers (cameras) see it from different perspectives, resulting in projective transformations. The transformations investigated here will be limited to the simpler subset of planar projective transformation, i.e., projective transformations of planar objects, which are described in section 2. Section 3 shows the derivation of the planar projective summation invariant. In section 4, this new feature is applied to a simulated camera network to show its usefulness for recognizing objects from different perspectives. Finally, section 5 contains a discussion of the results and future directions.

## 2. PLANAR PROJECTIVE TRANSFORMATIONS

The most general projective transformation is that from a 3D object to a 2D image, and is given by a 3 x 4 projection matrix, $P$. Thus, given a point on an object (in homogeneous coordinates), $\boldsymbol{x}$, the corresponding point in the image, $\boldsymbol{w}$, will be given by $\boldsymbol{w} = P\boldsymbol{x}$. If the object points are restricted to lie in the same plane, then the transformation is simplified to a 3 x 3 matrix. Thus, for planar projective transformations, the transformation is $\boldsymbol{w} = P^p \boldsymbol{x}$, or

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{w} \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{1}$$

where $[\tilde{x} \;\; \tilde{y} \;\; \tilde{w}]^T$ is the transformed point in homogeneous coordinates, and $P^p$ is the 3 x 3 planar projection matrix. If there are two separate planar projective transformations (corresponding to images from two cameras in different locations), $P^p$ and $Q^p$, and both transformations have non-zero determinants, the two image points corresponding to a point on the object are also related by a planar projective transformation $R^p$, where $R^p = P^p (Q^p)^{-1}$.

Each planar projective transformation only has eight degrees of freedom, so without loss of generality, $i$ can be set to 1 to simplify the computations. Equation (1) can be re-written as three equations as follows:

$$\tilde{x} = ax + by + c \tag{2}$$

$$\tilde{y} = dx + ey + f \tag{3}$$

$$\tilde{w} = gx + hy + 1 \tag{4}$$

To recover the Cartesian coordinates of this transformed point, each homogeneous coordinate is divided by $\tilde{w}$, e.g.,

$$\bar{x} = \tilde{x} / \tilde{w} = \frac{ax + by + c}{gx + hy + 1} \tag{5}$$

The nonlinearity caused by the presence of the variables x and y in the denominator of (5) will greatly complicate efforts to work with this transformation group to derive invariant features.

## 3. DERIVATION OF SUMMATION INVARIANTS

The derivation of summation invariants utilizes the method of moving frames developed by Cartan [11] and refined by Olver [12-13]. A detailed description of the method and its application to summation invariants is given in [14]. A brief description of the method is given here, with a focus on the parts that are especially relevant to the projective case.

Given a parameterized curve with points $(x[n], y[n])$, and a geometric transformation group $G$ acting on the curve, the group action is defined as:

$$g \circ (x[n], y[n]) = (\bar{x}[n], \bar{y}[n]), g \in G \tag{6}$$

where $(\bar{x}[n], \bar{y}[n])$ represents a transformed point in Cartesian coordinates. The moving frame method involves prolonging the transformation group action into the jet space, $J^m$ (consisting of the original coordinates and new coordinates formed by potentials up through order m), solving a normalization equation and plugging that solution into a higher order potential to obtain an invariant.

Potentials are defined as:

$$P_{i,j} = \sum_{n=1}^{N} x^i[n] \cdot y^j[n] \tag{7}$$

where $i + j = k$, with $k$ being the order of the potential. The formulation of potential is similar to the traditional definition of *image moment*. The moment for a discrete function, f(x), is:

$$m_i = \sum_x x^i \cdot f(x) \tag{8}$$

Substituting y = f(x), and parameterizing x and y by sample number n, $1 \le n \le N$, a parameterized definition of moment is:

$$m_i = \sum_{n=1}^{N} x^i[n] \cdot y[n] \tag{9}$$

From this definition, it is easy to verify that $m_i = P_{i,1}$. Similarly, $m_{i,j} = P_{i,j,1}$. Hence, the definition of potential is more general than the classical definition of image moment.

Then, the jet space is given by the coordinates:

$$\left(x[1], y[1], x[N], y[N], P_{(m)}\right) \tag{10}$$

where $P_{(m)}$ is all potentials up through order $m$. A transformed potential is defined as:

$$\bar{P}_{i,j} = \sum_{n=1}^{N} \bar{x}^i[n] \cdot \bar{y}^j[n] \tag{11}$$

For the case of the planar projective transformation group, an example of a transformed potential is:

$$\bar{P}_{1,0} = \sum_{n=1}^{N} \bar{x}[n] = \sum_{n=1}^{N} \left( \frac{\tilde{x}[n]}{\tilde{w}[n]} \right) = \sum_{n=1}^{N} \frac{ax + by + c}{gx + hy + 1} \tag{12}$$

where $\bar{P}_{1,0}$ represents the transformed potential in Cartesian coordinates. However, at this point in the standard procedure, the derivation attempt reaches an impasse. The technique of separating out the summation terms used for all the previous transformation groups (Euclidean and affine) cannot be applied because of the presence of the $x$ and $y$ terms in the denominator.

The key innovation that will allow the derivation of the new invariants to proceed is that of using homogenous coordinates for the transformed potentials instead of Cartesian. A transformed potential in homogenous coordinates is given by:

$$\tilde{P}_{i,j,k} = \sum_{n=1}^{N} \left( \tilde{x}^i[n] \cdot \tilde{y}^j[n] \cdot \tilde{w}^k[n] \right) \tag{13}$$

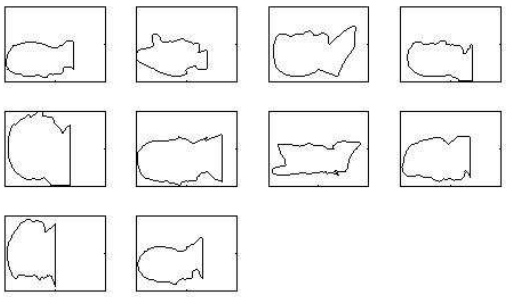

Fig. 1 - Original contour images.

An example potential for the planar projective transformation group is:

$$\tilde{P}_{1,0,0} = \sum_{n=1}^{N} \tilde{x}[n] = \sum_{n=1}^{N} (ax[n] + by[n] + c)$$
$$= aP_{1,0} + bP_{0,1} + cN \quad (14)$$

In this form of representation, the summation terms can now be separated out and a tractable formula derived. Although the transformed potential is in homogeneous coordinates, it can be expanded into a representation that is in terms of Cartesian coordinates, as seen in (14).

Similarly,

$$\tilde{P}_{0,1,0} = \sum_{n=1}^{N} \tilde{y}[n] = \sum_{n=1}^{N} (dx[n] + ey[n] + f) \quad (15)$$

$$\tilde{P}_{0,0,1} = \sum_{n=1}^{N} \tilde{w}[n] = \sum_{n=1}^{N} (gx[n] + hy[n] + 1) \quad (16)$$

The normalization equation used here is:

$$(\tilde{x}[1], \tilde{y}[1], \tilde{w}[1], \tilde{x}[N], \tilde{y}[N], \tilde{P}_{1,0,0}, \tilde{P}_{0,1,0}, \tilde{P}_{0,0,1})$$
$$= (0,0,1,1,0,1,1,1) \quad (17)$$

i.e., $\tilde{x}[1] = 0, \tilde{y}[1] = 0, ..., \tilde{P}_{0,0,1} = 1$.

Since there are eight degrees of freedom for this transformation group, the normalization equation requires eight equations. Using equation (17), solve equation (1) for {a,b,c,d,e,f,g,h}. Generating the invariant features is accomplished by applying these solutions to higher order transformed potentials (i.e., substituting the values for a, b, c, etc. into the equation for the transformed potentials). Thus, applying this procedure to $\tilde{P}_{0,2,0}$ yields the invariant:

$$\eta^{Pp}_{0,2,0} = (P_{20} \cdot y_N^2 - 2 \cdot P_{20} \cdot y_N \cdot y_1 + P_{20} \cdot y_1^2$$
$$+2 \cdot P_{11} \cdot y_N \cdot x_1 - 2 \cdot P_{11} \cdot y_N \cdot x_N - 2 \cdot P_{11} \cdot y_1 \cdot x_1$$
$$+ 2 \cdot P_{11} \cdot y_1 \cdot x_N - 2 \cdot P_{10} \cdot y_N^2 \cdot x_1 + 2 \cdot P_{10} \cdot y_N \cdot y_1 \cdot x_N$$
$$+ 2 \cdot P_{10} \cdot y_1 \cdot y_N \cdot x_1 - 2 \cdot P_{10} \cdot y_1^2 \cdot x_N + P_{02} \cdot x_1^2$$
$$- 2 \cdot P_{02} \cdot x_1 \cdot x_N + P_{02} \cdot x_N^2 - 2 \cdot P_{01} \cdot y_N \cdot x_1^2$$
$$+2 \cdot P_{01} \cdot x_1 \cdot y_1 \cdot x_N + 2 \cdot P_{01} \cdot x_N \cdot y_N \cdot x_1$$
$$- 2 \cdot P_{01} \cdot y_1 \cdot x_N^2 + N \cdot y_N^2 \cdot x_1^2 - 2 \cdot N \cdot y_N \cdot x_1 \cdot y_1 \cdot x_N + N \cdot y_1^2 \cdot x_N^2)$$
$$/(x_N \cdot P_{01} - x_N \cdot N \cdot y_1 - x_1 \cdot P_{01} - P_{10} \cdot y_N + P_{10} \cdot y_1 + x_1 \cdot N \cdot y_N)^2 \quad (18)$$

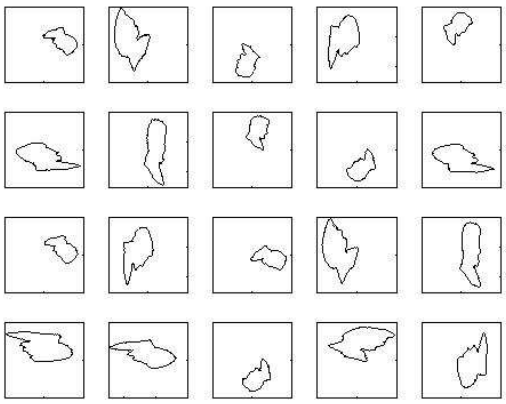

Fig. 2 – Twenty projective transformations of one original contour image.

where $x_1 = x[1]$, $x_N = x[N]$, etc. The same procedure was done for $\tilde{P}_{2,0,0}$ and $\tilde{P}_{1,1,0}$ to find $\eta^{Pp}_{2,0,0}$ and $\eta^{Pp}_{1,1,0}$.

The straight application of the summation invariant to an object will result in a single value. This is fine from the invariance perspective, but not good from the discrimination perspective. To improve the discrimination ability of these features, they can be applied in a semi-local manner. At each point of the object, the summation invariant is calculated over a subset of the surrounding points of the object. This yields an N-dimensional feature instead of a one-dimensional feature. The size of this window will determine the degree of localization of the feature.

## 4. APPLICATION TO CAMERA NETWORKS

In a typical camera network, it is desired to recognize an object that appears in multiple cameras. Each camera's image will represent a different projective transformation, with a typical network having a wide range of viewpoints. For effective recognition, a feature invariant to projective transformations is desired. This section provides simulation data demonstrating the application of the planar projective summation invariants derived here to object recognition in a camera network.

A database of ten images was created, consisting of the outlines of silhouettes of ten people. Each image was re-sampled to a size of 255 points. These images are shown in Fig. 1. To simulate a camera network, twenty random projective transformations were generated each representing a camera. The transformations were applied to each image, the resulting values were rounded to simulate camera quantization noise and post-processed to remove redundant points and re-sampled to 255 points, resulting in 200 images in the data set. The twenty transformed images from one original image are shown in Fig. 2.

The semi-local summation invariant values (numerator) were calculated for each image. The window size used was 51 points. Thus each contour had a 255 value feature vector, where each value was the 51 point summation invariant around that particular point in the image. The analysis of these results consisted of comparing each image with all the others and creating a 200 x 200 similarity matrix. The metric used was the normalized cross-correlation between the feature vectors,

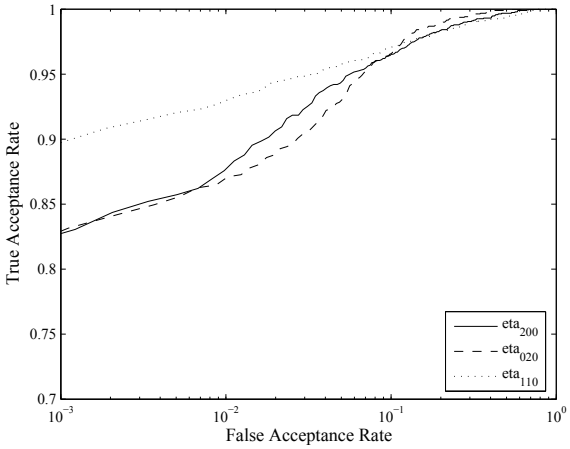


Fig. 3 – ROC performance for all three planar projective summation invariants on silhouette data.

$$\rho = \frac{\sum_{n=1}^{N}(\eta_1[n]\cdot\eta_2[n])}{\sqrt{\sum_{n=1}^{N}\eta_1^2[n]\cdot\sum_{n=1}^{N}\eta_2^2[n]}} \tag{19}$$

normalized to the range [0, 1.0]. The results are shown in Fig. 3 as an ROC graph comparing the three invariants to each other. Note that even though all three are invariant, their discrimination abilities are not equal.

The performance of these new features on this dataset was excellent, showing their effectiveness at discrimination across a wide range of projective transformations. Although the results are very good, the dataset was not extremely large, and the final judgment of the effectiveness of new features will require more extensive testing with larger and more challenging datasets.

A comparison was also done with the previous summation invariants derived under less general transformation groups (e.g., Euclidean, affine) using the standard comparison test defined in [14]. This test utilizes face recognition with the FRGC2.0 dataset. Although this application does not specifically require the projective transformation, it does show that these new features have similar discrimination capability as the previous ones. The comparison is shown in Table 1, where the numbers indicate the True Acceptance Rate at a False Acceptance Rate of 0.001.

Table 1 – Results of standard comparison test.

| | $\eta_{200}$ | $\eta_{110}$ | $\eta_{020}$ |
|---|---|---|---|
| 2D, Eucl., yN=0 | 0.581 | 0.719 | 0.681 |
| 2D, Affine | 0.684 | --- | --- |
| Planar proj. | 0.667 | 0.615 | 0.681 |

## 5. CONCLUSION

A new set of features was derived that are invariant to planar projective transformations. These features are an extension to the summation invariant family of features. The derivation of this new set of features overcame the challenges of projective transformations by performing the derivation in the homogeneous coordinate space. Application was made to the planar object recognition problem across a camera network, showing their effectiveness for this task.

Future directions for this work include comparison with other methods and simulations with larger datasets. Another area of interest is the extension to the projective transformation group, removing the planar object constraint and allowing for a broader application of the features.

## 6. REFERENCES

[1] E. Rivlin and I. Weiss, "Deformation invariants in Object Recognition," *Computer Vision and Image Understanding*, vol. 65, no. 1, pp. 95–108, 1997.

[2] Y. Avrithis, Y. Xirouhakis, and S. Kollias, "Affine-Invariant Curve Normalization for Object Shape Representation, Classification, and Retrieval," *Machine Vision and Applications*, vol. 13, pp. 80-94, 2001.

[3] G. Lei, "Recognition of Planar Objects in 3-D Space from Single Perspective Views Using Cross Ratio," Robotics and Automation, vol. 6, no. 4, pp. 432-437, 1990.

[4] K.W. Kwok, K.C. Lo, "Recognition of curved shapes using geometric invariants," *Fourth International Conference on Signal Processing (Cat. No.98TH8344*), 1998, pt. 2, p 1096-9 vol.2.

[5] S.C. Rajashekhar, V.P. Namboodiri, "Image retrieval based on projective invariance," *International Conference on Image Processing, ICIP,* 2004, p 405-408.

[6] C. E. Hann and M. S. Hickman, "Projective curvature and integral invariants," *Acta Applicandae Mathematicae*, vol. 74, no. 2, pp. 177–193, 2002.

[7] W. -Y. Lin, N. Boston, and Y. H. Hu, "Summation invariant and its application to shape recognition," in *Proceedings of ICASSP*, 2005, vol. V, pp. 205–208.

[8] W. -Y. Lin, K.-C. Wong, N. Boston, and Y. H. Hu, "3D Human Face Recognition Using Summation Invariants," in *Proceedings of ICASSP*, 2006, vol. II, pp. 341-344.

[9] W. -Y. Lin, "Robust Geometrically Invariant Features for 2D Shape Matching and 3D Face Recognition," Ph.D. dissertation, Dept. Elect. and Comp. Eng., Univ. of Wisconsin, Madison, WI, Aug. 2006.

[10] K.-C. Wong, W.-Y. Lin, Y. H. Hu, N. Boston, and X. Zhang, "Optimal Linear Combination of Facial Regions for Improving Identification Performance," to appear in IEEE Trans. on Systems, Man and Cybernetics, Part B: Cybernetics, submitted for publication.

[11] E. Cartan, "La methode du repere mobile, la theorie des groupes continus, et les espaces generalises," *Exposes de Geometrie*, no. 5, 1935.

[12] M. Fels and P. J. Olver, "Moving coframes: I. a practical algorithm," *Acta Applicandae Mathematicae*, vol. 51, no. 2, pp. 161 – 213, 1998.

[13] ——, "Moving coframes: II. regularization and theoretical foundations," *Acta Applicandae Mathematicae*, vol. 55, no. 2, pp. 127 – 208, 1999.

[14] K.R. Widder, W.-Y. Lin, N. Boston, Y.H. Hu, "From Moving Frames to Summation Invariants: Procedures, Properties and Application," Tech. Report ECE-07-05, Dept. Elec. Comp. Eng., Univ. of Wisconsin-Madison, Madison, WI, Sept. 2007.