On Representation Theory in Computer Vision Problems*

Amnon Shashua
School of Computer Science and Engineering
Hebrew University of Jerusalem
Jerusalem 91904, Israel
email: shashua@cs.huji.ac.il

Roy Meshulam
Department of Mathematics
The Technion
Haifa, Israel

Lior Wolf
School of Computer Science and Engineering
Hebrew University of Jerusalem
Jerusalem 91904, Israel

Anat Levin
School of Computer Science and Engineering
Hebrew University of Jerusalem
Jerusalem 91904, Israel

Gil Kalai
The Institute of Mathematics
Hebrew University of Jerusalem
Jerusalem 91904, Israel

Abstract

We introduce the following general question: Let V be a complex n-dimensional space and for $m \geq k$ consider the $GL(V)$-module $V(n,m,k) \subset V^{\otimes m}$ defined by

$$V(n,m,k) = \{ \quad v_1 \otimes \cdots \otimes v_m \in V^{\otimes m} \; : \; \dim \mathrm{Span}\{v_1, \ldots, v_m\} \leq k \; \} \, .$$

We would like to determine $\dim V(n,m,k)$ for any choice of $n, m \geq k$. This question appears in various disguises in computer vision problems where the constraints of a multi-linear problem occupy a low-dimensional subspace. We discuss two such problems: analysis of constraints in single view indexing functions (the 8-point shape tensor), and the analysis of the constraints in dynamic $\mathcal{P}^n \to \mathcal{P}^n$ alignments, i.e., where the point sets are allowed to move within a k-dimensional subspace while the n-dimensional space is being multiply projected (multiple views) onto copies of the m-dimensional space. We then derive the solution to the general problem using tools from representation theory.

1 Introduction

Multilinear constraints in computer vision applications are of growing interest in Structure for Motion (SFM), Indexing and Graphics. Many of the applications where multiple mea-

*The reference to this manuscript is “Technical Report 2002-44, Leibniz Center for Research, School of Computer Science and Eng., the Hebrew University of Jerusalem.”

surements are involved — like multiple-view geometry of static and dynamic scenes, indexing functions into 3D data-sets, separation of various attribute/modalities such as "content" and "style" — have a multilinear form. As a result, a growing amount of work has been published on the various aspects of those algebraic functions and their applications — see [10, 6] for the recent summary of various multi-linear maps and their associated tensors.

In this paper we raise a general question and demonstrate its relevance to the current research in multilinearity in computer vision. The questions takes the following form: Let V be a complex n-dimensional space and for $m \geq k$ consider the $GL(V)$-module $V(n,m,k) \subset V^{\otimes m}$ defined by

$$V(n,m,k) = \{ \quad v_1 \otimes \cdots \otimes v_m \in V^{\otimes m} \ : \ \dim \mathrm{Span}\{v_1, \ldots, v_m\} \leq k \ \} \ .$$

We would like to determine $\dim V(n,m,k)$ for any choice of $n, m \geq k$. We will show that this question appears in a one disguised form or another in a number of vision problems and, for example, focus on two of those problems: (i) analysis of constraints in single view indexing functions (the 8-point shape tensor), and (ii) the analysis of the constraints in dynamic $\mathcal{P}^n \rightarrow \mathcal{P}^n$ mappings, i.e., where the point sets are allowed to move within a k-dimensional subspace while the n-dimensional space is being multiply projected (multiple views) onto copies of the m-dimensional space.

We then derive the solution to the general problem using tools from representation theory. We will describe the general notations in the next section (and provide a brief primer on representation theory in the appendix), followed by the detailed description of the two problems mentioned above and the way the are mapped to the question of $\dim V(n,m,k)$, and followed by the derivation of the structure and dimension of the $GL(V)$ module $V(n,m,k)$ by counting irreducibles followed with examples of its application to some instances of dynamic $\mathcal{P}^n \rightarrow \mathcal{P}^n$ mappings.

2 Notations

We will describe below the notations and symbols we will be using later in the paper. A brief account of the relevant facts concerning the representation theory of the general linear group can be found in the Appendix.

Let V be a finite n-dimensional vector space over the complex numbers, and let the group of automorphisms of V denoted by $GL(V)$. We denote the exterior powers of V by $\wedge^m V$ and the symmetric powers by $\mathrm{Sym}^m V$. A partition of m is denoted by $\lambda = (\lambda_1, ..., \lambda_k)$ such that $\lambda_1 \geq ... \geq \lambda_k \geq 1$ and $\sum \lambda_i = m$. A partition is represented by its Young diagram (or "shape") which consists of k left aligned rows of boxes with λ_i boxes in row i. We denote by μ_i the number of terms in λ that are greater than or equal to i, and $\mu = (\mu_1, ..., \mu_r)$ is called the conjugate partition of λ.

We denote by $\mathcal{T}_\lambda$ the set of standard tableaux on λ and f_λ the number of standard tableaux on λ,

$$f_\lambda = \frac{m!}{\prod_{(i,j)} h_{ij}}$$

where $h_{ij} = \lambda_i + \mu_j - i - j + 1$ is called the "hook length" of a box in position (i,j), and the product of the hook-lengths is over all boxes of the diagram. We denote by $d_\lambda(n)$ the number of semi-standard tableaux:

$$d_\lambda(n) = \prod_{(i,j)} \frac{n - i + j}{h_{ij}}.$$

Let t be a tableau on λ (a numbering of the boxes of the diagram) and let $P(t)$ denote the group of all permutations $\sigma \in S_m$ which permute only the rows of t. Similarly, let $Q(t)$

denote the group of permutations that preserve the columns of t. Let a_t, b_t be two elements in the group algebra $\mathbb{C}S_m$ defined as:

$$a_t = \sum_{g \in P(t)} g \ , \ b_t = \sum_{g \in Q(t)} \mathrm{sgn}(g) g,$$

and we denote Schur's Module by $S_t(V) = V^{\otimes m} \cdot a_t \cdot b_t$.

3 The 8-point Shape Tensor Problem

In this section we will make the connection between the question of $\dim V(n,m,k)$ and a riddle regarding the internal structure of the 8-point shape tensor. Shape tensors were first introduced in [2, 16, 3] with the basic idea that single-view invariants of a 3D scene can be obtained by algebraically eliminating the viewing position (camera) parameters given a sufficient number of points. Later, the same analysis was conducted in a reduced (but practical in vision applications) setting where a reference plane is identified in advance [11, 12, 5, 4] — which is the case we will focus on here.

The problem setting is as follows. Let $P_i = (X_i, Y_i, Z_i, W_i)^\top \in \mathcal{P}^3$, $i = 1, ..., 8$, denote 8 points in 3D projective space and let M be a 3×4 projection matrix, thus $p_i \cong MP_i$ where $p_i \in \mathcal{P}^2$ be the corresponding image points in the 2D projective plane. We wish to algebraically eliminate the camera parameters (matrix M) by having a sufficient number of points. This could be done succinctly if we first make a change of basis: Let the coplanar points be denoted by $P_1, ..., P_4$ with the coordinates $(1,0,0,0), (0,1,0,0), (0,0,1,0), (1,1,1,0)$ which is appropriate when $P_1, ..., P_4$ are indeed coplanar. Let the image undergo a projective change of coordinates such that the corresponding points $p1, ..., p4$ be assigned $e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1), e_4 = (1,1,1)$, respectively. Given this setup the camera matrix M contains only 4 non-vanishing entries:

$$M = \begin{bmatrix} \delta & 0 & 0 & \alpha \\ 0 & \delta & 0 & \beta \\ 0 & 0 & \delta & \gamma \end{bmatrix}$$

Let $\hat{M} = (\alpha, \beta, \gamma, \delta) \in \mathcal{P}^3$ be a point (representing the camera) and let $\hat{P}_i$ be the projection matrix:

$$\hat{P} = \begin{bmatrix} W_i & 0 & 0 & X_i \\ 0 & W_i & 0 & Y_i \\ 0 & 0 & W_i & Z_i \end{bmatrix}$$

And as in the general case we have the duality $p_i \cong MP_i = \hat{P}_i \hat{M}$ where the role of the motion (the camera) and shape have been switched. Let l_i, l'_i be two distinct lines passing through the image point p_i, i.e., $p_i^\top l_i = 0$ and $p_i^\top l'_i = 0$, and therefore we have $l_i^\top \hat{P}_i \hat{M} = 0$ and $l_i'^\top \hat{P}_i \hat{M} = 0$. For $i = 5, ..., 8$ we have therefore $E\hat{M} = 0$ where:

$$E = \begin{bmatrix} l_5^\top \hat{P}_5 \\ \cdot \\ l_8^\top \hat{P}_8 \\ l_5'^\top \hat{P}_5 \\ \cdot \\ l_8'^\top \hat{P}_8 \end{bmatrix} \tag{1}$$

Therefore the determinant of any 4 rows of E must vanish. The choice of the 4 rows can include 2 points, 3 points, or 4 points (on top of the 4 basis points $P_1, ..., P_4$) and each such choice determines a multilinear constraint whose coefficients are arranged in a tensor. The 8-point tensor is when 4 points are chosen: by choosing one row from each point we obtain a vanishing determinant involving 4 points which provides 16 constraints (per view) $l_i^5 l_j^6 l_k^7 l_t^8 \mathcal{Q}^{ijkt} = 0$ for the 81 coefficients of the tensor $\mathcal{Q}^{ijkt}$. The indices i, j, k, l follow the covariant-contravariant notations (upper index represents points, lower represent lines) and follow the summation convention (contraction) $u^i v_i = u^1 v_1 + u^2 v_2 + ... + u^n v_n$. The tensor contains 81 coefficients, however, they satisfy internal ("synthetic" borrowing from [10]) linear constraint. *Exactly how many is an open problem which we will show boils down to the question of* $\dim V(n, m, k)$.

Since $P_1, ..., P_4$ are coplanar we have the constraint $P_i^\top n = 0$, $i = 1, ..., 4$ and, due to our choice of coordinates, $n = (0, 0, 0, 1)^T$. Consider the family of camera matrices $M = un^\top$ for all choices of $u = (u_1, u_2, u_3)^\top$. In other words, the 4'th column of M consists of the arbitrary vector u and all other entries vanish. Thus we have that MP either vanishes or is equal to u (up to scale) for all P. Let l_i, l'_i be lines through u, therefore

$$\begin{aligned} l_i^\top M_j P &= l_i^\top \hat{P} \hat{M}_j = 0 \\ l_i'^\top M_j P &= l_i'^\top \hat{P} \hat{M}_j = 0 \end{aligned}$$

for all points P, and dually for all projection matrices $\hat{P}$. Therefore the 4×4 determinants of E vanish regardless of $\hat{P}_i$. We have a single $3 \times 3 \times 3 \times 3$ tensor $\mathcal{Q}^{ijkt}$ responsible for the 16 quadlinear constraints $l_i^5 l_j^6 l_k^7 l_t^8 \mathcal{Q}^{ijkt} = 0$ (we have a choice of 2 lines for each point, thus 16 constraints). From the discussion above, the four lines contracted by the tensor are all coincident with the arbitrary point u. Therefore, *the question is what is the dimension of the set of constraints* $l_i^5 l_j^6 l_k^7 l_t^8 \mathcal{Q}^{ijkt} = 0$ *where the lines are arbitrary but form a 2-dimensional subspace?*

Recall the definition of $V(n, m, k)$ and set $n = 3, m = 4, k = 2$:

$$V(3, 4, 2) = \{v_1 \otimes v_2 \otimes v_3 \otimes v_4 | \dim \mathrm{Span}\{v_1, ..., v_4\} \leq 2\}$$

where $v_1, ..., v_4$ are vectors in R^3. Our question regarding the number of synthetic constraints is equivalent to the question of *what is the dimension of* $V(3, 4, 2)$?

4 Dynamic $\mathcal{P}^n \rightarrow \mathcal{P}^n$ Mappings

Consider a configuration of points in $Q_i \in \mathcal{P}^{n-1}$, $i = 1, ..., q$ undergoing a projective mapping $Q_i \rightarrow Q'_i$. Then it is well known that $Q'_i \cong AQ_i$ where $A \in GL(n)$ is some invertible $n \times n$ matrix. However, consider the following "complication" where each point Q_i may *change* its position up to a k-dimensional subspace ($k = 1$ means that Q_i is fixed, $k = 2$ means that Q_i may change its position along some line in $\mathcal{P}^n$, and so forth), and we are given $m > 2$ observations $Q_i^{(j)}$ where $j = 1, ..., m$. In other words, the observations $Q_i^{(j)}$ are generated by a combination of "global" (unknown) transformations $A_i \in GL(n)$ and "local" (unknown) movements within (unknown) subspaces of dimension up to $k < m$. The task is to recover the global transformations A_i from the observations.

The definition above is a generalization of particular cases which were introduced in the past under the name of "dynamic" SFM, or SFM of multiply moving points, and the relevant literature includes [1, 15, 19, 13, 17, 8, 14, 9, 18]. For instance, [15] consider the case where $n = 3$ (points Q_i belong to the 2D projective plane), $m = 3$ and $k = 2$. In other words, a configuration of coplanar points are viewed by a moving camera and the points move along arbitrary straight lines ($k = 2$) or stay fixed ("static", $k = 1$) while the camera changes positions. It was shown there that the image observations (across three views) satisfy a $3 \times 3 \times 3$ tensorial constraint, where in the case where all points are moving along along lines, 26 observations are sufficient for a unique solution to the tensor, when all points are static (without being labeled as such) then those observations fill a 10 dimensional subspace (thus at least 16 points should be dynamic for a unique solution form observations).

In a later paper [19] the case of "dynamic 3D to 3D" alignment was introduced, where $n = 4, m = 3, k = 2$. In that case, the observations are governed by a $4 \times 4 \times 4$ tensor, where the observations from moving points fill a 60-dimensional space (thus there 4 tensors satisfying the constraints), and static points fill a 20-dimensional space.

Among the various aspects of those tensors, one important aspect is the counting of necessary constraints for a solution. Some of those counting issues, even in the particular low dimension examples given above, are not obvious. The matter becomes fairly subtle when dealing with the general dynamic $\mathcal{P}^n \rightarrow \mathcal{P}^n$ mappings where the issue of counting constraints is an open problem.

We observe that since tensor products commute with linear transformations, the issue of dimension counting is independent of the matrices $A_i \in GL(n)$. Therefore, the general problem of counting the constraints of a dynamic $\mathcal{P}^{n-1} \rightarrow \mathcal{P}^{n-1}$ mapping is isomorphic to the question of $\dim V(n, m, k)$, where in this case $n \geq m \geq k$.

When we compute the constraints of dynamic mappings we have other limitations which are not described in [15, 19] and can be also described in the $V(n, m, k)$ framework. For example, in the case of dynamic $\mathcal{P}^2 \rightarrow \mathcal{P}^2$ alignment the collection of measurements arising from triplets of matching points must span the 2D plane. We may ask what is the largest number of collinear points allowed? (which beyond that the solution becomes degenerate). In other words, the question is how many points moving on the same striaght line path will generate linearly independent constraints. The answer is $\dim V(2, 3, 2)$ — note that $n = 2$ because the effective dimension of the vector space is 2 even though the points are in defined in the 2D projective plane (i.e., $n = 3$). Likewise, in the case of dynamic $\mathcal{P}^2 \rightarrow \mathcal{P}^2$ alignment the maximal number of points allowed on a single line is also $\dim V(2, 3, 2)$ — and out of these points $\dim V(2, 3, 1)$ static points will give us linearly independent constraints (in both cases).

From the examples above we have that $\dim V(3, 3, 2) = 26$ and $\dim V(4, 3, 2) = 60$ (point moving along straight line paths) and $\dim V(3, 3, 1) = 10$ and $\dim V(4, 3, 1) = 20$ (static points) for the 2D and 3D cases, respectively.

In the following section we analyze the structure of $V(n, m, k)$ and as a result determine $\dim V(n, m, k)$ for any choice of $n, m \geq k$.

5 The Structure of $V(n, m, k)$

So far we have presented two (unrelated) Vision problems which are isomorphic to the $\dim V(n, m, k)$ question. We will provide below the statement and proof about the structure of $V(n, m, k)$. The statement appears very similar to the classic result (see Appendix) of decomposing of $V^{\otimes m}$ into irreducible $GL(V)$-modules:

$$V^{\otimes m} = \bigoplus_{\lambda \vdash m} \bigoplus_{t \in \mathcal{T}_\lambda} S_t(V),$$

with the difference that *not all* diagrams are included — only those diagrams λ for which $\lambda_{k+1} = 0$.

Claim 1

$$V(n, m, k) = \bigoplus_{\lambda_{k+1}=0} S_\lambda(V)^{\oplus f_\lambda} .$$

In particular

$$\dim V(n, m, k) = \sum_{\lambda_{k+1}=0} f_\lambda s_\lambda .$$

Proof: suppose $\lambda \vdash m$ and $\lambda_{k+1} = 0$. Let t be the tableau given by $t(i,j) = \sum_{l=1}^{i-1} \lambda_l + j$. Noting that $V(n,r,1) = \mathrm{Sym}^r V$ it follows that

$$\begin{aligned} V^{\otimes m} \cdot a_t &= \mathrm{Sym}^{\lambda_1} V \otimes \cdots \otimes \mathrm{Sym}^{\lambda_k} V \\ &= V(n,\lambda_1,1) \otimes \cdots \otimes V(n,\lambda_k,1) \subset V(n,m,k)\ . \end{aligned}$$

Therefore,

$$S_t(V) = V^{\otimes m} \cdot a_T \cdot b_T \subset V(n,m,k) \cdot b_T \subset V(n,m,k)$$

hence,

$$\bigoplus_{\lambda_{k+1}=0} S_\lambda(V)^{\oplus f_\lambda} \subset V(n,m,k).$$

To show the other direction let $(\cdot,\cdot)$ be a hermitian form on V and let the induced form on $V^{\otimes m}$ be given by

$$(u_1 \otimes \cdots \otimes u_m, v_1 \otimes \cdots \otimes v_m) = \prod_{i=1}^{m} (u_i, v_i)\ .$$

Note that

$$\begin{aligned} &(u_1 \wedge \cdots \wedge u_m, v_1 \otimes \cdots \otimes v_m) \\ &= \frac{1}{m!}(u_1 \wedge \cdots \wedge u_m, v_1 \wedge \cdots \wedge v_m) \\ &= \frac{1}{m!} \det[(u_i, v_j)]_{i,j=1}^m\ . \end{aligned}$$

Let $\lambda \vdash m$ with $\lambda_{k+1} \neq 0$, then the conjugate partition $\mu = (\mu_1 \geq \mu_2 \geq \ldots \geq \mu_t)$ satisfies $\mu_1 \geq k+1$. Let $l_j = \sum_{r=1}^{j} \mu_r$ and let t be the tableau given by $t(i,j) = l_{j-1} + i$. Then

$$\begin{aligned} S_t(V) &= V^{\otimes m} \cdot a_t \cdot b_t \subset V^{\otimes m} \cdot b_t \\ &= \wedge^{\mu_1} V \otimes \cdots \otimes \wedge^{\mu_l} V\ . \end{aligned}$$

Suppose now that $v_1, \ldots, v_m \in V^{\otimes m}$ satisfy $\dim \mathrm{Span}\{v_1, \ldots, v_m\} \leq k$. Then $v_1 \wedge \cdots \wedge v_{\mu_1} = 0$ therefore for any $u_1, \ldots, u_m \in V$

$$\begin{aligned} &((u_1 \otimes \cdots \otimes u_m) \cdot b_T, v_1 \otimes \cdots \otimes v_m) = \\ &\prod_{r=1}^{l} \frac{1}{\mu_r!} \Big(\bigwedge_{i=l_{r-1}+1}^{l_r} u_i, \bigwedge_{i=l_{r-1}+1}^{l_r} v_i \Big) = 0\ . \end{aligned}$$

It follows that $V(n,m,k)$ is orthogonal to

$$\bigoplus_{\lambda_{k+1}\neq 0} \mathcal{S}_\lambda(V)^{\oplus f_\lambda}$$

hence,

$$\dim V(n,m,k) \leq \dim \bigoplus_{\lambda_{k+1}=0} \mathcal{S}_\lambda(V)^{\oplus f_\lambda}\ .$$

□

Claim 1 can be used to give explicit formulas for $\dim V(n,m,k)$ when either k or $m-k$ are small. In the later case we write

$$\dim V(n,m,k) = n^m - \sum_{\lambda_{k+1}\neq 0} f_\lambda d_\lambda(n)$$

and note that the partitions of m with $\lambda_{k+1} \neq 0$ correspond to all partitions of all numbers up to $m-k-1$.

5.1 Examples

To calculate $\dim V(n,m,m-1)$ note that only $\lambda=(1^m)$ must be excluded, thus:

$$f_{(1^m)}=1\ ,\ d_{(1^m)}(n)=\binom{n}{m}$$

hence,

$$\dim V(n,m,m-1)=n^m-\binom{n}{m}.$$

To calculate $\dim V(n,m,m-2)$ we must exclude, in addition to the above, the partition $(2,1^{m-2})$, thus:

$$f_{(2,1^{m-2})}=m-1\ ,\ d_{(2,1^{m-2})}(n)=(m-1)\binom{n+1}{m}$$

hence,

$$\dim V(n,m,m-2)=n^m-[\binom{n}{m}+(m-1)^2\binom{n+1}{m}].$$

To calculate $\dim V(n,m,m-3)$ we must exclude, in addition to the above, the partitions $(3,1^{m-3})$ and $(2^2,1^{m-4})$, thus:

$$f_{(3,1^{m-3})}=\binom{m-1}{2}\ ,\ d_{(3,1^{m-3})}(n)=\binom{m-1}{2}\binom{n+2}{m}$$

$$f_{(2^2,1^{m-4})}=\frac{m(m-3)}{2},$$

$$d_{(2^2,1^{m-4})}(n)=\frac{(m-3)n}{2}\binom{n+1}{m-1}$$

Hence,

$$\dim V(n,m,m-3)=n^m-[\binom{n}{m}+(m-1)^2\binom{n+1}{m}+$$

$$\binom{m-1}{2}^2\binom{n+2}{m}+\frac{m(m-3)^2n}{4}\binom{n+1}{m-1}].$$

With these in mind, we can easily resolve the first of the open problems which is the number of synthetic constraints of the 8-point shape tensor with 4 coplanar points. We have seen that the answer is $\dim V((3,4,2)$:

$$\dim V((3,4,2)=\sum_{\lambda_3=\lambda_4=0}f_\lambda d_\lambda,$$

where $\lambda=(\lambda_1,...,\lambda_4)$, is a partition of 4, i.e., $\lambda_1\geq\lambda_2\geq\lambda_2\geq\lambda_4$ and $\sum_i\lambda_i=4$. We have therefore only three partitions which satisfy $\lambda_3=\lambda_4=0$: $\lambda=(4),(2,2),(3,1)$ to consider. Thus, $f_{(4)}=1,d_{(4)}=15,f_{(2,2)}=2,d_{(2,2)}=6,f_{(3,1)}=3$ and $d_{(3,1)}=15$. Therefore, $\dim V(3,4,2)=15+12+45=72$.

We can also verify the special cases of dynamic $\mathcal{P}^2\rightarrow\mathcal{P}^2$ and $\mathcal{P}^3\rightarrow\mathcal{P}^3$ by substituting the values of n,m,k in the formulas above. For example: $\dim V(3,3,2)=27-1=26$ and $\dim V(4,3,2)=64-4=60$ (point moving along straight line paths) and $\dim V(3,3,1)=27-(1+4\cdot4)=10$ and $\dim V(4,3,1)=64-(4+4\cdot10)=20$ (static points). Also $\dim V(2,3,2)=8-0=8$ points moving along one line path out of which up to $\dim V(2,3,1)=8-[0+4]=4$ are static points on this line will give us linearly independent constraints.

6 Summary

We have shown that certain non-obvious counting problems exist in SFM literature such as the number of synthetic constraints of the 8-point shape tensor with 4 coplanar points, and the number of constraints necessary for the general dynamic $\mathcal{P}^n \rightarrow \mathcal{P}^n$ alignment problem — and in general in problems where the constraints of a multi-linear problem occupy a low-dimensional subspace.

We have shown that a certain general question lies at the heart of those counting problems: Let V be a complex n-dimensional space and for $m \geq k$ consider the $GL(V)$-module $V(n,m,k) \subset V^{\otimes m}$ defined by

$$V(n,m,k) = \{ \quad v_1 \otimes \cdots \otimes v_m \in V^{\otimes m} \;: \\ \dim \mathrm{Span}\{v_1, \ldots, v_m\} \leq k \;\} \,.$$

We would like to determine $\dim V(n,m,k)$ for any choice of $n, m \geq k$. Thus, for instance we showed that $\dim V(3,4,2)$ is the number of synthetic constraints of the 8-point shape tensor with 4 coplanar points, and $\dim V(n,m,k)$ stands for the number of constraints of a $\mathcal{P}^{n-1} \rightarrow \mathcal{P}^{n-1}$ alignment problem with m mappings and where the points move in a $k-1$ dimensional subspaces.

We have then shown that the questions of $\dim V(n,m,k)$ is naturally addressed in the context of representation theory by counting the irreducibles of $V^{\otimes m}$ over a *subset* of diagrams of the partition of m.

It is worthwhile noting that representation theory tools have not been used so far in the computer vision literature, thus the fact that such problems exist in the context of vision tasks suggest that some familiarity with these kind of tools would bear fruits also in future research.

References

[1] S. Avidan and A. Shashua. Trajectory triangulation: 3D reconstruction of moving points from a monocular image sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(4):348–357, 2000.

[2] S. Carlsson. Duality of reconstruction and positioning from projective views. In *Proceedings of the workshop on Scene Representations*, Cambridge, MA., June 1995.

[3] S. Carlsson and D. Weinshall. Dual computation of projective shape and camera positions from multiple images. *International Journal of Computer Vision*, 27(3), 1998.

[4] C. Rother and S. Carlsson. Linear Multi View Reconstruction and Camera Recovery. In *Proceedings of the International Conference on Computer Vision*, Vancouver, Canada, July 2001.

[5] A. Criminisi, I. Reid, and A. Zisserman. Duality, rigidity and planar parallax. In *Proceedings of the European Conference on Computer Vision*, Frieburg, Germany, 1998. Springer, LNCS 1407.

[6] O. Faugeras and Q.T. Luong with contributions from T. Papadopoulo. *The geometry of multiple images* MIT Press, 2001.

[7] W.Fulton and J.Harris. *Representation Theory: a First Course.* Springer-Verlag, 1991.

[8] M. Han and T. Kanade. Reconstruction of a Scene with Multiple Linearly Moving Objects. In *Proc. of Computer Vision and Pattern Recognition*, June, 2000.

[9] M. Han and T. Kanade. Multiple Motion Scene Reconstruction from Uncalibrated Views. In *Proceedings of the International Conference on Computer Vision*, Vancouver, Canada, July 2001.

[10] R.I. Hartley and A. Zisserman. *Multiple View Geometry*. Cambridge University Press, 2000.

[11] M. Irani and P. Anandan. Parallax geometry of pairs of points for 3D scene analysis. In *Proceedings of the European Conference on Computer Vision*, LNCS 1064, pages 17–30, Cambridge, UK, April 1996. Springer-Verlag.

[12] M. Irani, P. Anandan, and D. Weinshall. From reference frames to reference planes: Multiview parallax geometry and applications. In *Proceedings of the European Conference on Computer Vision*, Frieburg, Germany, 1998. Springer, LNCS 1407.

[13] R.A. Manning and C.R. Dyer. Interpolating view and scene motion by dynamic view morphing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 388–394, Fort Collins, Co., June 1999.

[14] D. Segal and A. Shashua 3d reconstruction from tangent-of-sight measurements of a moving object seen from a moving camera. In *Proceedings of the European Conference on Computer Vision*, Dublin, Ireland, June 2000.

[15] A. Shashua and Lior Wolf. Homography tensors: On algebraic entities that represent three views of static or moving planar points. In *Proceedings of the European Conference on Computer Vision*, Dublin, Ireland, June 2000.

[16] D. Weinshall, M. Werman, and A. Shashua. Duality of multi-point and multi-frame geometry: Fundamental shape matrices and tensors. In *Proceedings of the European Conference on Computer Vision*, LNCS 1065, pages 217–227, Cambridge, UK, April 1996. Springer-Verlag.

[17] Y. Wexler and A.Shashua. On the synthesis of dynamic scenes from reference views. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, South Carolina, June 2000.

[18] Lior Wolf and A. Shashua. On projection matrices $\mathcal{P}^k \longrightarrow \mathcal{P}^2$, $k = 3, \cdots, 6$, and their applications in computer vision. In *Proceedings of the International Conference on Computer Vision*, Vancouver, Canada, July 2001.

[19] Lior Wolf, A. Shashua, and Y. Wexler. Join tensors: on 3d-to-3d alignment of dynamic sets. In *Proceedings of the International Conference on Pattern Recognition*, Barcelona, Spain, September 2000.

A A Representation Theory Digest

In this section we briefly recall some relevant facts concerning the representation theory of the general linear group. For a thorough introduction see [7].

Let V be a finite n-dimensional vector space over the complex numbers. The collection of invertible $n \times n$ matrices is denoted by $GL(n)$ which is the group of automorphisms of V denoted by $GL(V)$. The vector space $V^{\otimes m}$ (m-fold tensor product) is spanned by decomposable tensors of the form $v_1 \otimes \cdots \otimes v_m$, where the vectors v_i are in V. Hence the dimension of $V^{\otimes m}$ is n^m. The vector space $V^{\oplus m}$ is the m-fold direct sum of V, thus is of dimension nm.

The *exterior* powers $\wedge^m V$ of V, $n \geq m$, is the vector space spanned by the $m \times m$ minors of the $n \times m$ matrix $[v_1, ..., v_m]$ where the vectors v_i are in V. Hence the dimension of $\wedge^m V$ is $\binom{n}{m}$. The exterior powers are the images of the map $V^{\times m} \rightarrow V^{\otimes m}$ given by

$$(v_1, \cdots, v_m) \rightarrow \sum_{\sigma \in S_m} \text{sgn}(\sigma) v_{\sigma(1)} \otimes, \cdots, v_{\sigma(m)}$$

where S_m denotes the symmetric group (of *permutations* of m letters).

The *symmetric powers* $\mathrm{Sym}^m V$ are the images of the map $V^{\times m} \to V^{\otimes m}$ given by

$$(v_1, \cdots, v_m) \to \sum_{\sigma \in S_m} v_{\sigma(1)} \otimes, \cdots, v_{\sigma(m)}$$

Hence the vector space $\mathrm{Sym}^m V$ is of dimension $\binom{n+m-1}{m}$. Note that,

$$V \otimes V = \mathrm{Sym}^2 V \oplus \wedge^2 V$$

with the appropriate dimension: $n^2 = \binom{n+1}{2} + \binom{n}{2}$. This decomposition into irreducibles (see later) is not true for $V^{\otimes m}$, $m > 2$. The remainder of this section is devoted to the necessary notation for representing $V^{\otimes m}$ as a decomposition of irreducibles.

A *representation* of a group G on a complex finite dimensional space U is a homomorphism G to $GL(U)$ - the group of linear automorphisms of U. The action of $g \in G$ on $u \in U$ is denoted by $g \cdot u$. The $G-$module U is *irreducible* if it contains no non-trivial $G-$invariant subspaces. Any finite dimensional representation of a compact group G can be decomposed as a direct sum of irreducible representations. This basic property called *complete reducibility* also holds for all holomorphic representations of the general linear group $GL(V)$.

The main focus of this paper is the space

$$V(n,m,k) = \mathrm{Span}\{v_1 \otimes \cdots \otimes v_m \in V^{\otimes m} \; : \; \dim \mathrm{Span}\{v_1, \ldots, v_m\} \leq k \; \} \; .$$

Since $V(n,m,k)$ is invariant under the $GL(V)$ action given by $g \cdot v_1 \otimes \cdots \otimes v_m = g(v_1) \otimes \cdots \otimes g(v_m)$ it is natural to study its structure by decomposing it into irreducible $GL(V)$-modules.

The description of the finite dimensional irreducible representations (irreps) of $GL(V)$ depends on the Combinatorics of partitions and Young diagrams which we now describe: A *partition* of m is an ordered set $\lambda = (\lambda_1, ..., \lambda_k)$ such that $\lambda_1 \geq ... \geq \lambda_k \geq 1$ and $\sum \lambda_i = m$. A partition is represented by its *Young diagram* (also called *shape*) which consists of k left aligned rows of boxes with λ_i boxes in row i. The *conjugate partition* $\mu = (\mu_1, ..., \mu_r)$ to a partition λ is defined by interchanging rows and columns in the Young diagram — or without reference to the diagram, μ_i is the number of terms in λ that are greater than or equal to i.

An assignment of the numbers $\{1, ..., m\}$ to each of the boxes of the diagram of λ, one number to each box, is called a *tableau*. A tableau in which all the rows and columns of the diagram are increasing is called a *standard tableau*. We denote by f_λ the number of standard tableaux on λ, i.e., the number of ways to fill the young diagram of λ with the numbers from 1 to m, such that all rows and columns are increasing. Let (i,j) denote the coordinates of the boxes of the diagram where $i = 1, .., k$ denotes the row number and j denotes the column number, i.e., $j = 1, ..., \lambda_i$ in the i'th row. The *hook length* h_{ij} of a box at position (i,j) in the diagram is the number of boxes directly below plus the number of boxes to the right plus 1 (without reference to the diagram, $h_{ij} = \lambda_i + \mu_j - i - j + 1$). Then,

$$f_\lambda = \frac{m!}{\prod_{(i,j)} h_{ij}}$$

where the product of the hook-lengths is over all boxes of the diagram. We denote by $d_\lambda(n)$ the number of *semi-standard tableaux* which is the number of ways to fill the diagram with the numbers from 1 to n, such that all rows are non-decreasing and all columns are increasing. We have:

$$d_\lambda(n) = \prod_{(i,j)} \frac{n - i + j}{h_{ij}}.$$

Let S_m denote the symmetric group on $\{1, \ldots, m\}$. The *group algebra* $\mathbb{C}S_m$ is the algebra spanned by the elements of S_m

$$\mathbb{C}G = \{\sum_{\sigma \in S_m} \alpha_\sigma \sigma \mid \alpha_\sigma \in \mathbb{C}\}$$

where addition and multiplication are defined as follows:

$$\alpha(\sum_{\sigma \in S_m} \alpha_\sigma \sigma) + \beta(\sum_{\sigma \in S_m} \beta_\sigma \sigma) = \sum_{\sigma \in S_m} (\alpha\alpha_\sigma + \beta\beta_\sigma)\sigma$$

and

$$(\sum_{\sigma \in S_m} \alpha_\sigma \sigma)(\sum_{\tau \in S_m} \beta_\tau \tau) = \sum_{g \in S_m} (\sum_{g = \sigma\tau} \alpha_\sigma \beta_\tau) g$$

for $\alpha, \beta, \alpha_\sigma, \beta_\sigma \in \mathbb{C}$.

Let t be a tableau on λ (a numbering of the boxes of the diagram) and let $P(t)$ denote the group of all permutations $\sigma \in S_m$ which permute only the rows of t. Similarly, let $Q(t)$ denote the group of permutations that preserve the columns of t. Let a_t, b_t be two elements in the group algebra $\mathbb{C}S_m$ defined as:

$$a_t = \sum_{g \in P(t)} g \ , \ b_t = \sum_{g \in Q(t)} \mathrm{sgn}(g) g.$$

The group algebra $\mathbb{C}S_m$ acts on $V^{\otimes m}$ on the right by permuting factors, i.e., $(v_1 \otimes \cdots \otimes v_m) \cdot \sigma = v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(m)}$. For a general shape λ and a tableau t on λ the image of a_t, $V^{\otimes m} \cdot a_t$, is the subspace:

$$V^{\otimes m} \cdot a_t = \mathrm{Sym}^{\lambda_1} V \otimes \cdots \otimes \mathrm{Sym}^{\lambda_k} V \subset V^{\otimes m}$$

and the image of b_t is

$$V^{\otimes m} \cdot b_t = \wedge^{\mu_1} V \otimes \cdots \otimes \wedge^{\mu_r} V \subset V^{\otimes m}$$

where μ is the conjugate partition to λ. The *Young symmetrizer* is defined by $c_t = a_t \cdot b_t \in \mathbb{C}S_m$. The image of the Young symmetrizer

$$S_t(V) = V^{\otimes m} \cdot c_t$$

is the *Schur Module* associated to t and is an irreducible $GL(V)$- module. The isomorphism type of $S_t(V)$ depends only on the shape λ so we may write $S_t(V) = S_\lambda(V)$. It turns out that all the polynomial irreps of $GL(V)$ are of the form $S_\lambda(V)$ for some m and a partition $\lambda \vdash m$.

Let $\mathcal{T}_\lambda$ denote the set of standard tableaux on λ then the direct sum decomposition of $V^{\otimes m}$ into irreducible $GL(V)$-modules is given by

$$V^{\otimes m} = \bigoplus_{\lambda \vdash m} \bigoplus_{t \in \mathcal{T}_\lambda} S_t(V) \cong$$

$$\bigoplus_{\lambda \vdash m} \mathcal{S}_\lambda(V)^{\oplus f_\lambda} .$$

Since $d_\lambda(n) = \dim \mathcal{S}_\lambda(V)$ it follows that

$$\dim V^{\otimes m} = n^m = \sum_{\lambda \vdash m} d_\lambda(n) f_\lambda.$$

For example, consider $n = m = 3$, i.e., $V \otimes V \otimes V$ where $\dim V = 3$. There are three possible partitions λ of 3 — these are $(3), (1,1,1)$ and $(2,1)$. From the above, $S_{(3)}(V) = \mathrm{Sym}^3 V$ and $S_{(1,1,1)} V = \wedge^3 V$. There are two, $f_{(2,1)} = 2$, standard tableaux for $\lambda = (2,1)$ and these are 123 and 132 (numbering of boxes left to right and top to bottom). There are eight, $d_{(2,1)}(3) = 8$, semi-standard tableaux which are: $112, 113, 122$,123, $132, 133$,223 and 233. We have the decomposition:

$$V \otimes V \otimes V = \mathrm{Sym}^3 V \oplus \wedge^3 V \oplus (S_{(2,1)} V)^{\oplus 2}$$

with the appropriate dimensions: $27 = 10 + 1 + (8 + 8)$.