SYMBOLIC MODELSFOR POSTURES RECOGNITION OF A THREE FINGERED ARTIFICIAL HAND

M. Gonzalo–Tasis, R. Pellón, D. Sánchez and J. Finat
MOBiVA- Group of Advanced Visualization, Dept. of Computer Science, Univ. Of Valladolid, Spain
marga@infor.uva.es, rpellon@tid.es, dsanchez@gmv.es, jfinat@nava.tel.uva.es

Working Group V/4

KEY WORDS: Image Understanding, Model-based processing

ABSTRACT

Hand posture and gesture recognition integrates sensorial fusion, supervised model-based learning and motion planning. Data acquisition of changing postures in known environments requires the development of intelligent systems based on flexible models which can be updated maintaining properties (incidence and order) of knuckles regarding to phalanxes.

An essential characteristic of this model is the modularity that allow us: to identify hand postures based on geometric information, use implicit anatomic and physiological hierarchy of an anthropomorphic three-fingered hand and finally identify postures with neural fields. In this paper we have selected an approach based in visual inputs only. These visual inputs correspond to a symbolic representation of the skeleton.

Classification and interpretation process are controlled in terms of symbolic hybrid models able to integrate geometric information (meaningful for free obstacle navigation of the artificial hand) with neural fields. Geometric information is relative to node positions (some of them represent knuckles) and segments (representing visible boundaries of phalanxes). Neural fields provide autonomous decision mechanism from acquisition and processing of non-linear activation/inhibition processes depending on stimula.

1 INTRODUCTION

Traditional algebro-geometric methods are based in a minimal set of points necessaries to identify postures, (seven points [LK95] or six points in [MC97]). Nevertheless, their efficiency from a mathematical viewpoint is very sensitive to partial occlusions due to grasping or self-occlusions that arise from motions of an articulated hand. Instead, we work with redundant information acquired from a higher number of control points; resulting systems are overdetermined and it is necessary optimize these systems to solve them. The optimization process is not easy, it is essential to work with flexible data structures that can be able to integrate mobile points and segments.

Our approach is simple; it is based on a low-level identification of grouped segments verifying metric conditions for pose estimation and incidence conditions for invariants detection, that reinforces postures identification.

Posture recognition for an artificial model can be understood as a mapping between a subset of meaningful features of the articulated hand and some model features stored in some database. Hence, the process begins with a descending high-level supervised geometric model simulated with OPEN/GL. Next, we use a low-level processing to extract meaningful geometric features to be compared (points, segments, or both) and their incidence properties. Finally, we identify the nearest mechanical events corresponding to flexioned / extended phalanxes that act as possible attractors for evolving postures relative to each finger. This information is symbolically stored following a simplified mechanical version of aspect graphs, where each simple mechanical event corresponds to an interchange between an extension and a flexion for each knuckle of each finger.

The states of the system are generated in symbolic representations by extracting information from discontinuities of pixel functions. Also, with aprioristic knowledge based in model, it can be obtained invariant geometric and topologic descriptions. This qualitative information about states is utilized as coarse reference models working as attractors and thus, it must be invariant for the acquired representation.

Symbolic models are not enough accurate at first sight, but they provide a low-level support allowing integrate different kinds of information.

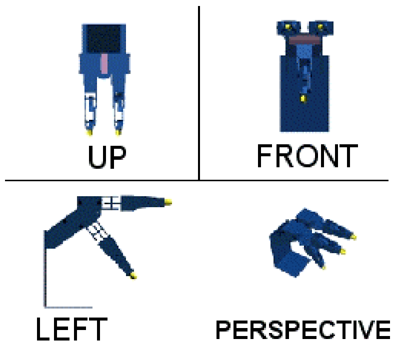
Most popular models are variants of the well-known Kohonen's Self-Organizing Maps algorithm [Ko97]. Parametrised SOM provide differentiable manifolds as models, whereas Local Linear Maps (LLM) associate to each neuron a locally valid linear mapping (see [Ri97] and references therein).

Overimposed structure given by the tangent bundle (with its dual the cotangent bundle or phase space) allow us to connect vertically both variants in terms of transition functions. These functions express changes of references to connect the same visual or mechanical events from different aspect graphs; or alternately, to construct paths between adjacent vertices in symbolic representations.

2 IMAGE INTERPRETATION

2.1 Hand Images Generation

Our goal is to identify postures from a simulated Stanford/JPL three-fingered artificial hand generated by Open/GL. OpenGL is versatile, adaptable and portable.



We have inspired in a hand model well-known with a complex architecture and functionality (3 DOF for each finger with cylindrical and piecewise linear components) similar to an anthropomorphic hand.

We have develop a 3D model and with the OpenGL virtual camera we obtained monocameral successive views of the hand.

Each view, that is a bidimensional projection of the scene, is preprocessed converting in a 256-gray level, and subtracting the background image of the hand image

2.2 Low-Level Processing

From this solid modeling, SUSAN (Smallest Univalue Segment Assimilating Nucleus) allows us to process in low-level each image, extracting geometric characteristics as points and segments, that verify incidence conditions. These characteristics correspond to visible parts of the shape of 3D articulated hand. SUSAN generates a coarse estimation of visible boundaries and meaningful points modeled and grouped; additionally we could extract geometrical information from graphical evaluation of the depth of meaningful points in terms of ray tracing.

Thus, we obtain a set of segments that bounds the shape of the simulated artificial hand and a set of corners (see junction detections in [GF98b]) identified in terms of fast variations of curvature from the boundary shape.

3 POSTURES RECOGNITION

Multiple junctions appearing in the knuckles are not stable; depending on the posture, they can be evolving from T (extension) to an *arrow* ↑ (flexion) or vice versa. Hence, identification of junctions would require perform tracking and grouping in a simultaneous way, by taking care with partial or self-occlusions (see [FG98]), but this approach is too expensive in time.

Then, instead of tracking multiple junctions we extract regions that furthermore the multiple character of some junctions, each region contributes only to a double ordinary corner. In fact, this is procedure more stable to small perturbations and noise corrupting. In this way, each T-junction corresponding to an extension of a knuckle is captured as two L-junctions, one for each region. The same operation is carried out with other triple junctions.

We found a set of pair of parallel segments (modulus some threshold depending on robotic or human hand) taking those who have minimal separation distance. This makes easier the construction of a virtual skeleton of the hand. Virtual skeleton is the key for posture identification in terms of incidence conditions.

To fix ideas, the i-th finger has three DOF (in the configuration space) represented by a column vector is $\theta^i = (\theta_1^i, \theta_2^i, \theta_3^i)^T$ to which we associate $s(\theta^i) = (s_1^i, s_2^i, s_3^i)^T$ where $s_j^i = sin(\theta_j^i)$ with $1 \leq j \leq 3$. In particular the entries of each vertex are binary i.e. extended knuckle is 0 and flexioned knuckle is 1; additionally we can round off values of s_j^i in order to know the nearest extremal posture for each finger.

Binary values for each finger can be represented as vertices of a 3D cube that we associate to each finger. Similarly, each edge of this cube does correspond to replace a 0 by means an 1; thus, it represents an evolving posture from an extension to a flexion for two adjacent segments (representing adjacent phalanxes) in a vertex (common knuckle).

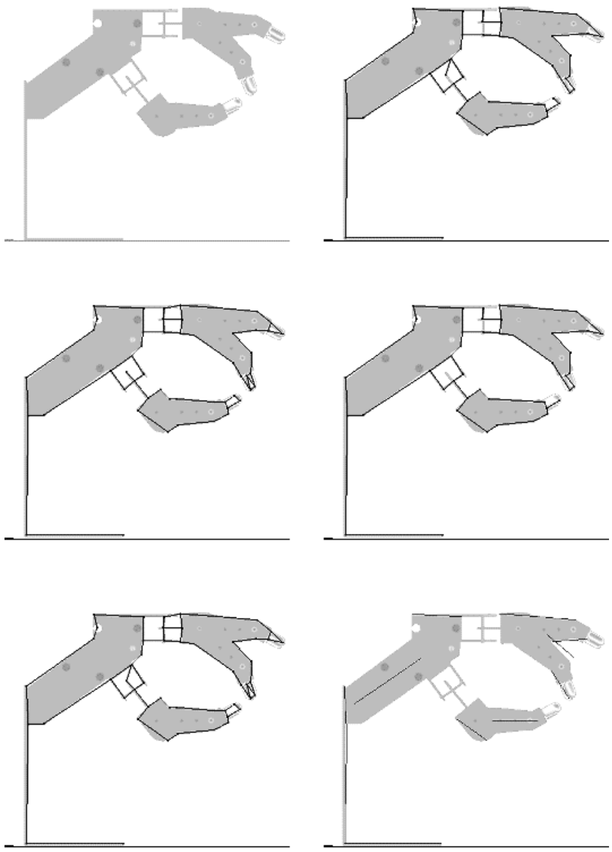


Fig2: Successive process of an image

The exact position for each finger must be acquired and learned along edges connecting adjacent vertices, depending on parameters controlling kinematic and dynamic restrictions (these effects have not been considered here). From the kinematic viewpoint, each vertex in each cube corresponding to i-th finger acts as a possible attractor between competitive postures, in such way that we can label the posture in terms of the nearest attractor. Then, elementary dynamical aspects for evolving postures are formulated in terms of competitive neural fields for each finger [SFP99].

Nevertheless their appearance, this formulation gives a discrete low-level non-deterministic representation, which is dynamically preserved in unions of 3D cubes for each finger. This algebraic representation of each geometric posture as a 3x3-matrix for each finger is matched in a 3x3k-matrix corresponding to coordinated tasks controlled by a parallel processor. Evolving postures (gestures) are modeled as paths, whose binary matrices (obtained by rounding inputs) change only one datum each time they arrive to some vertex. Due to parallel character of this updating for each 3x3-matrix, we can have several simultaneous mechanical events represented as a simultaneous change in each box. This presentation has several advantages from the geometrical viewpoint based on some elementary applications of Grassmannians as mathematical tool ([FG98]), and from the block dynamics acting on the columns of matrix, instead of an individual activation/inhibition phenomenon.

4 INTERPRETATION OF IMAGES AS POSTURES

Images interpretation concerns to a medium level in the *Recognition Phase*. Interpretation is performed with reference to coarse geometric and kinematic models. Geometric models are based on symbolic representations of an articulated hand in terms of a graph guided by the largest segments obtained in each image. Euclidean version of this method is strongly dependent of the viewpoint, but its symbolic representation by means an adjacency graphs is efficient and robust w.r.t. small perturbations, and this symbolic approach does not need complicated evaluation of invariant properties w.r.t. rigid and scale transformations due to strong constraints about articulated mechanisms. Furthermore, it allows us to simplify postures analysis due to robustness of tracking segments instead of tracking control points. In this way, in this model, it is possible to integrate control based on position and control based on trajectories, without making a cumbersome analysis of kinematics.

Thus, in this article we have concentrated our attention in posture recognition based on geometric aspects, in detriment of another aspects related to the tracking and interaction with environment.

Firstly, we select the parallel segments with greatest width corresponding to the palm of the hand. We assign a variable weight to parallel segments, which is proportional to relative orientation of the camera w.r.t. to the hand. A correct identification of the palm is crucial for the rest of process. After identification of the palm, we evaluate characteristics of segments associated to the fingers, which are common between them.

Fingertips and base points (corresponding to first knuckles) are located in regions where we have many corners. Selection procedure depends on largest segments connecting such corners. In this case, we need also to assign a variable weight allowing us to make an automatic selection of segments corresponding to fingers meaningful for each posture. These weights have been obtained in an experimental way. After this selection, one can verify that making use of only a small number of segments performs posture recognition: usually two or three segments are sufficient (in the worst case, we need eight segments to perform this identification without error).

Coarse kinematic models are obtained following hypothesis propagation between standard geometric models, which act as possible attractors for evolving postures. Determination of the hand kinematics is a very hard mathematical problem, due to our partial knowledge about non-linear effects for each finger, about modeling of distributed processes involving to different fingers and coordination problems between fingers to perform a concrete task. Additional dynamical aspects related to stiffness and compliance that are crucial for coordination have not been considered here.

Traditional postures tracking are an approach very expensive and very dependent of the specific model we are looking for ([RK94). For, in most presentation one needs a very concrete model, well-specified states for the model and some knowledge about kinematics connecting different well-known states. In this scheme, one must characterise geometric features, which are meaningful for each posture, to measure these features, to estimate states and track their spatio-temporal evolution. Furthermore, if we wish to interact with the artificial hand, one must to make a 3D reconstruction arising from some stereo vision system.

Next challenge is to deduce inverse kinematics from postures learning in order to improve man-machine interaction. This goal requires to transfer visual information to motor representation, under partial or incomplete information, i.e. without having an exact knowledge of dynamical equations controlling the motion, but only some partial qualitative data arising from effects observed by neural fields. These neural fields act by means stimula on very few neurones, by selecting the preferred orientation and activation/inhibition level for each task on each knuckle.

Multivector representing this orientation in the configuration space gives the posture up to scale by means a (3x3)-matrix corresponding to the $sin(\theta_{ij})$, where θ_{ij} represents nine relative angles. Relative angles are given as differences between angles corresponding to consecutive phalanxes in virtual knuckles (points where segments symbolising phalanxes are intersecting). To improve accuracy, it must to add geometric information about the DOF of articulated hand and allowed movements of the fingers. This representation (as first-order differences) is especially well behaved to a discrete approach to kinematic questions, which can be easily adapted to Neurodynamical Models [Ha94].

The description of neural fields in the working space is not easy; for, we would need to specify euclidean information about control points where we must apply neural fields. Instead, we describe several basic neural fields depending on the agonist /antagonist situation (corresponding to flexion/extension postures of involved muscles). We have a specific neural field for each finger (not only for each neuron, because we have functionally similar neurons in different fingers), a codebook for different joints allowing us to discriminate between signals corresponding to different articulations appearing in each finger. Activation/inhibition mechanisms of neural fields depend on the task and the proximity relations w.r.t. obstacles presenting in the scene.

5 TRACKING WITH NEURAL FIELDS

Neural Fields can be understood as some kind of dynamical systems biologically inspired on a geometrical support. They are associated to an artificial model of the configuration C or the working space W. Existence of a natural projection $\Pi : C \to W$ allow us transfer these *Neural Fields* between both spaces, and display natural feedback for perception-action cycle.

Dynamical aspects appear in Neural Networks early in the 60s, because the prominent role-played by the time in neuronal dynamics. Nonlinearity in the generation, processing and transmission of impulses, are the sources of a lot of sophisticated mathematical approaches with biomechanical and neurophysiological basis. However, we can simplify adopting here a discrete model based on first-order difference equations for model acquisition and recognition of static postures, by excluding another more complex dynamic phenomena linked to gestures (evolving postures along the time). The mathematical description of above nonlinear phenomena would require higher-order equations (reaction-diffusion equations) and their computational implementation is far from the scope of this paper.

Our hybrid approach uses symbolic models based on matricial representation which contain information about configuration space C as starting point to initiate the network. Features incorporation arising from sensors is one permanent challenge, because very often it would be desirable have our disposal horizontal/vertical regional patterns for activation/inhibition. Our implementation follows a piecewise linear approach (similar to [DBJ98] for continuous models).

In an independent work [SFP99] we have implemented a multilayer perceptron artificial neural network where we have introduced the Little dynamics. This choice is justified by its parallel character (the updating of neurons is carried out simultaneously), its simple mathematical formulation (it is given by first-order equations) and its behavior (synchronous way in simplest deterministic models without delays). Using systems depending on adjustable parameters depending on model diminishes its initial deterministic character (PSOM: Parametrised Self-Organised Maps). In addition, there can emerge a nonlinear dynamics from a linear domain (transient from laminar flow to turbulence, e.g. in Fluid Mechanics).

Parallelism condition is meaningful to perform an independent information treatment and control relative to each finger, and to correct errors in an independent way, before using the supervisory network controlling global aspects relative to complex tasks. In addition, coupling and decoupling processes (cooperative and competitive behavior) are easily implemented without using sophisticated simulations. The only information we are considering for control is the difference between actual and desired position-orientation parameters for meaningful segments for posture recognition. This information is more robust w.r.t. to partial (self)occlusions and noise than those based on nodes corresponding to knuckles, and it is able of supporting activation/inhibition muscular effects (measurable in eletrophysiological terms). In addition, activation/inhibition elementary patterns are easily modeled in matricial terms, which makes easier their extension to massive systems.

6 EXPERIMENTS

We have used a six-stage recognition hierarchy from a simulated monocameral and monochrome input image. We generate and display in a simultaneous way several high-resolution images corresponding to three orientations of the same object in order to select (simulate in real experiments) the best location for a (virtual camera) allowing us to identify the current posture. Instead of using a grid or planar lattice, we identify relative orientation of the palm of the hand from geometric characteristics of the largest closed region (eventually a segment for certain views).

REFERENCES

[Ah95] S. Ahmad "A usable real-time 3D hand tracker" in 28th Asilomar Conference on Signals, Systems and Computers, IEEE computer Society Press, 1995.

[DBJ98] P. Dahm, C. Burckhoff And F. Joublin: "A Neural Field Approach for Robot Motion Control", in IEEE Int. Conf. SMC'98 (San Diego, CA, October, 1998), 3460-3465.

[DS94] J. Davis And M. Shah: "Recognizing Hand Gestures", in J.O.Eklundh (ed.): Computer Vision ECCV'94 Springer-Verlag, LNCS 800 (1994), 331-340.

[FG98] J. Finat And M. Gonzalo-Tasis: "Fast recognition of postures for a simplified three-fingered hand", IEEE Int. Conf. SMC’98 (San Diego, CA), October (1998).3466-3470

[FL98] J. Finat And J. Lopez-Coronado: "An Hybrid Model for the Hand Preconfiguration in Rehabilitation Grasping Tasks", IEEE Int. Conf. SMC’98 (San Diego, CA), October (1998), 3448-3453.

[GF98b] M. Gonzalo-Tasis And J. Finat: "Towards a Coarse Real-Time Stereo Fusion based on Triple Junctions", SIBGRAPI'98, Rio de Janeiro, October, 1998.

[GF98a] M. Gonzalo-Tasis And J. Finat: "A hybrid approach for a real time stereo fusion based on triple junctions", Proc. of the IV Jornadas de Informática, Las Palmas de Gran Canaria Spain, Junio 1998.

[HE97] P. A. Harling And A. D .N. Edwards, eds: Progress in Gestural Interaction, Springer-Verlag, 1997.

[Hay94] Haykin: Neural Networks, IEEE Press, 1994.

[Ho99] R.M.Hord: Understanding Parallel Supercomputing , IEEE Press, 1999.

[Ko97] T.Kohonen: Self-organizing maps, Springer Series in Information Sciences, Springer-Verlag, 1997.

[KBM96] J. Kosecka, R. Bajcsy And M. Mintz: "Control of Visually Guided Behaviors", in C.Brown and D.Terzopoulos (eds): Real-time Computer Vision Cambridge University Press (1996), 189-207.

[LK95] J.Lee And Y.Kunii: "Model-based analysis of hand posture", in IEEE Computer Graphics and Applications (1995), 77-86.

[MC97] R.J.Millar And G.F.Crawford: "A Mathematical Model for Hand-Shape Analysis" in [HE97], 235-245.

[Pe94] P.Peretto: An Introduction to the Modeling of Neural Networks, Cambridge Univ. Press, reprinted 1994.

[PS99] R.Perez-Pellon And D.Sanchez-Pardo: Modelos Simbólicos para el Reconocimiento de Posturas de una Mano Artificial de Tres Dedos, Master Thesis, Univ of Valladolid, Enero, 1999.

[RK94] J.M.Rehg And T.Kanade: "Visual Tracking of High DOF Articulated Structures: an application to Human Hand Tracking", in J.O.Eklundh (ed): ECCV'94 Stockholm, LNCS 801,Springer-Verlag (1994), 35-46.

[Ri97] H.Ritter: "Neural Network Approaches for Perception and Action", in G.Sommer and J.K.Koendering, eds: International Workshop, AFPAC'97, Kiel/, LNCS 1315 (1997), 356-378.

[SA97] K.Surendro And Y.Anzai: "Non-rigid Object Recognition using Principal Component Analysis and Geometric Hashing", in G.Sommer, K.Daniilidis and J.Pauli, eds: 7th Intl Conf, CAIP'97, Kiel, September, 1997, LNCS 1296, Springer-Verlag 50-57.

[Su] SUSAN: (Smallest Univalue Segment Assimilating Nucleus) by Steve Smith/Oxford University (http://www.robots.ox.ac.uk:80/steve/susan/index.html).

[UFSK95] Y. Uno, N. Fukumura, R. Suzuki And M. Kawato: "A Computational Model for the Recognizing Objects and Planning Hand Shape in Grasping Movements", IEEE Neural Networks, Vol.8, No.6 (1995), 839-851.