Hand tracking by 3D model registration

Hocine Ouhaddi, Patrick Horain

Unité de projets ARTEMIS, Département Signal et Images
Institut National des Télécommunications,
9, rue Charles Fourier,
91011 Evry cedex - France

tel : +33 (0)1.60.76.47.02, +33 (0)1.60.76.47.48
fax : +33 (0)1.60.76.44.34
e-mail : hocine.ouhaddi@int-evry.fr, patrick.horain@int-evry.fr
http://www.int-evry.fr/~hand

Abstract

We address the issue of 3D hand gesture analysis by monoscopic vision without body markers. A 3D articulated model is registered with images sequences. We compare several registration evaluation functions (edge distance, non-overlapping surface) and optimisation methods (Levenberg-Marquardt, downhill simplex and Powell). Biomechanical constraints are integrated into the minimisation algorithm to constrain registration to realistic postures. Results on image sequences are presented. Potential application include hand gesture acquisition and human machine interface.
Keywords : gesture analysis, hand tracking, model registration, cost function, optimisation, biomechanical constraints.

1 Introduction

Gestures are a natural way of communication for people. They also can be used for human-machine communication [1]. Unfortunately, datagloves are expensive, fragile and limit natural interaction. Therefore, gesture analysis by artificial vision is a topic of interest.
There are two different approaches to this problem : 2D aspect analysis [2] or 3D model based analysis [3]. 2D methods can recognise a limited number of configurations, generally after a learning procedure ; 3D analysis methods replace *a priori* knowledge on gesture with *a priori* knowledge on the observed object, namely : a hand. Most importantly, the 3D approach allows full 3D modelling of gestures. Thus it opens the path to humanoid animation and gesture-based human-machine interaction. For these reasons, we preferred the 3D model-based approach for gesture acquisition.

In this paper, we start recalling briefly our hand model and its adjustment on an image with respect to anthropometrical constraints. The second part presents briefly the state of the art of gesture analysis (hand, shoulder or body) by model registration. The third part presents our approach of gesture modelling. Finally, results are presented on real sequences of hand gestures.

2 Hand modelling

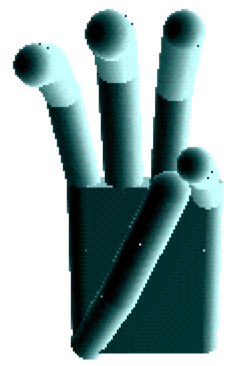


Figure 1 : model of the hand.

The hand model (figure 1) used in our experiments is made of 2 layers. The first layer describes the articulated model as a graph of geometric objects with colour attributes, translations, rotations and groups of these primitives. It implements all 26 degrees of freedom (DOF). Six are global (translations and rotations) for the wrist. The 4 DOF of each finger are : abduction/adduction and flexion/extension for the first articulation, flexion/extension for the second and third articulation. Biomechanical constraints allow reduce the number of DOF to 21.

The bounding box of each articulated component specifies its size according to Wagner's anthropometric constraints on the morphological measures [4]. The second layer is a description of the shape of articulated components.

In general, this model does not match hand morphology of a given operator. The model parameters need to be adjusted in advance to ensure robustness of hand tracking. Our approach, of unsupervised type, uses a single image of an open hand.

After extracting the silhouette and identifying the different parts of the hand, morphology parameters of fingers and palm are estimated from the dimensions of the corresponding regions. Anthropometric constraints allow to check the consistency of the estimated morphological measures as well as to estimate length of phalanxes.

3 Gesture analysis by model registration : state of the art

Mochimaru and Yamazaki [5] and then Kuch and Huang [6] register an articulated 3D model on images as a cost function minimisation. Their procedure consists in local variations of the freedom degrees. Namely, they apply small modifications to the value of each degree of freedom, and repeat that operation in the same direction if the cost decreases, or in the opposite direction otherwise. Degrees of freedom are processed finger by finger hierarchically starting from the first phalanx. Although this basic descent algorithm seems to be efficient in practice, its many parameters have to be chosen empirically.

Ohya and Kishino [7] register a human body model by a genetic algorithm that mimics natural selection in a population of the state vectors. The reported results require hundreds of iterations, which are computationally expensive.

Gavrila and Davis [8] optimise a human body model registration after discretisation of the set of parameters and hierarchical searching in a tree representing possible

configurations. Shimada *et al.* [9] register a hand model by generating a set of discrete candidate poses, both from the current image features and from the previous positions in the sequence. They use statistic dependence rules between successive configurations. Candidates are progressively selected as the image sequence is processed. Discretisation of the parameter space requires a compromise between precision and computation costs.
In the next part we consider 3D hand model registration in the continuous parameter space, with minimisation methods that do not require empirically tuned parameters.

4 Cost function for registration

Registration consists in finding out the model parameters that minimise the difference between the projected hand model and the corresponding features extracted from images. For this comparison, we have considered the surface differences between the silhouettes, or the sum of distances between their contours.

4.1 Surface cost function

We define this cost function as :

$$F = \frac{\mathrm{NOS}(\mathrm{modelSil}, \mathrm{imageSil}) + \mathrm{NOS}(\mathrm{imageSil}, \mathrm{modelSil})}{\mathrm{Surf}(\mathrm{modelSil}) + \mathrm{Surf}(\mathrm{imageSil})}$$

where the not overlapped surface $\mathrm{NOS}(\mathrm{sil1},\mathrm{sil2})$ is the number of pixels of the silhouette $\mathrm{sil1}$ that do not belong to $\mathrm{sil2}$. $\mathrm{Surf}(\mathrm{sil})$ is the surface of the silhouette. This cost function is to be minimised by some algorithm described later. Because it is locally flat when object do not overlap, the minimisation process must start from an initial position that shows some overlapping. That condition can usually be fulfilled by matching the centres of gravity of the projected model and image silhouette.

4.2 Contours distance

The cost function to be minimised is the sum of the distances between the silhouette pixels of the projected model and of the extracted silhouette normalised by the contours length. Distance from pixels to contours are computed by a Chamfer algorithm [10,11].

5 Optimisation methods

Some of the optimisation methods require evaluations of the cost function to be minimised, while some other also require evaluations of the derivatives of this function.

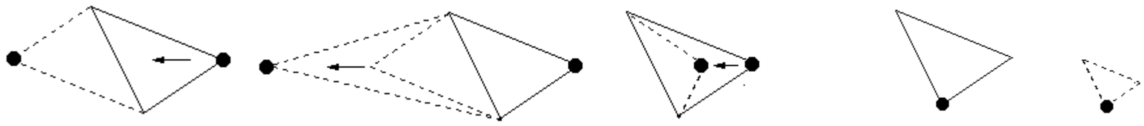
5.1 The Levenberg-Marquardt method

The Levenberg-Marquadt method is a steepest gradient algorithm using partial derivatives of the error function. It uses a parameter in order to circumvent rank

deficiency of the jacobian matrix (used to calculate correction vector) and to best control the iterative optimisation process.
The Levenberg-Marquardt has proved to be a very efficient non-linear optimisation method [12].

5.2 The downhill simplex method

Cost function derivatives are difficult to estimate. We have also considered optimisation methods that do not require any derivation. The "downhill simplex" method was proposed by Nelder and Mead [13]. Let n be the number of parameters. The first stage of the algorithm consists in generating a simplex of $n+1$ points in R^n starting from an initial solution. Geometrical transformations are then iteratively applied to the simplex : symmetries, expansions, contractions and multi-dimensional contraction (figure 2), on condition that the new simplex is not degenerated. These transformations allow to adapt the simplex to the local topography of the cost function. The transformation process stops when the relative variation of the cost function at the simplex vertices gets close to the computer precision. Our implementation is derived from the paper of Press *et al.* [14]. Other implementations can be found in [15].



(a) symmetry (b) expansion (c) contraction 1D (d) contraction ND

Figure 2 : Geometrical transformations of the simplex in R^2

5.3 The Powell method

There are many different methods for minimising a mono-dimensional cost function without applying derivatives, such as the dichotomy algorithm, the Fibonnacci's method or the golden section method [14]. To minimise a multidimensional function, it is possible to repeat mono-dimensional minimisation successively in the direction of each parameter axis. If the cost function contains a valley that is not oriented along a parameter axis, this method requires a lot of small displacements. The Powell's method rather minimises the cost function by changing search direction to valley direction, so speeding up the optimisation process [14].

6 Choice of optimisation method and cost function

We have compared the combinations of the error functions and minimisation methods, except the Levenberg-Marquardt method on non-overlapping surface because derivatives cannot be computed easily.
The tests have been done with various synthesis and natural sequences. Our results follow :

- Downhill simplex optimisation on the silhouette surface allows to track rigid translations of hand and abduction/adduction and flexion/extension of fingers, even when initialisation is far from optimum. However the algorithm may converge to a *bad* local minimum. This can be worked out by restarting the whole optimisation process with a larger initial simplex.
- The Powell optimisation on the silhouette surface gives the same results, but requires more cost functions evaluations. This is because Powell method uses a mono-dimensional minimisation procedure that performs more evaluations of the cost function [14].
- Registration with contour distances appears to be very sensitive to both model initialisation and image noise. Optimisation methods then fail to converge to a good minimum. This is due to the fact that the distance transform achieves an implicit matching between close contours. Consequently, the non overlapping surface appears to give better results than distances to contours. This meets the conclusion in [16] that region information stabilises tracking.

Therefore non overlapping surface with Downhill simplex appears to be the most efficient of these approaches.

7 Biomechanical constraints

Biomechanics address body movement. Fingers movements are governed by static and dynamic biomechanical constraints [6]. The former introduce limiting intervals of joint angles. The latter deals with interdependence between different articulations DOF. Among the most important :

- Interdependence between abduction/adduction and flexion/extension of a finger :

$$\theta_{a/a}^{MCP\ finger} = \theta_{a/a\max}^{MCP\ finger} - \frac{\theta_{a/a\max}^{MCP\ finger}}{\theta_{f/e\max}^{Mcp\ finger}} \times \theta_{f/e}^{MCP\ finger}$$

The higher the flexion/extension, the lower the abduction/adduction (figure 3a).

- Interdependence between flexion/extension of adjacent fingers :

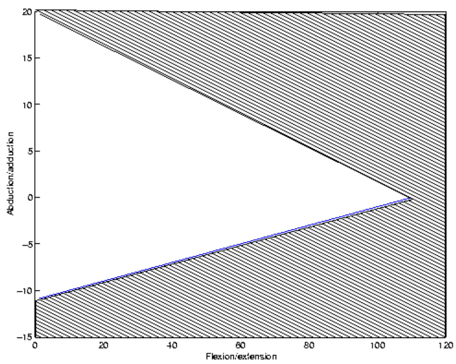
$$\min\begin{pmatrix} \theta_{f/e}^{MCP\ Index} {}^{\circ}+ 25^{\circ} \\ \theta_{f/e}^{MCP\ Ring} + 20^{\circ} \\ \theta_{f/e\max}^{MCP\ Middle} \end{pmatrix} \geq \ \theta_{f/e}^{MCP\ Middle} \quad \geq \max\begin{pmatrix} \theta_{f/e}^{MCP\ Index} - 25^{\circ} \\ \theta_{f/e}^{MCP\ Ring} - 45^{\circ} \\ \theta_{f/e\min}^{MCP\ Middle} \end{pmatrix}$$

This constraint means that the flexion of the index or the ring causes the flexion of the middle (figure 3b).

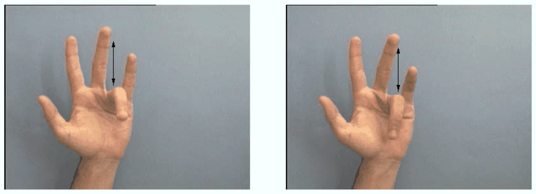
8 Downhill simplex optimisation with biomechanical constraints

A priori knowledge about realistic and non-realistic configurations of the hand is introduced in the optimisation procedure as set of inequalities to be satisfied. This defines a convex domain in the parameter space.

The biomechanical constraints can be integrated with the downhill simplex algorithm by bounding the simplex transformations inside this domain. Because the constraints



(a) Interdependence between abduction/adduction and flexion/extension of a finger constraint

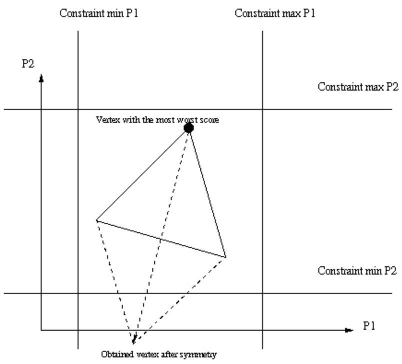


b) Interdependence between ring and middle

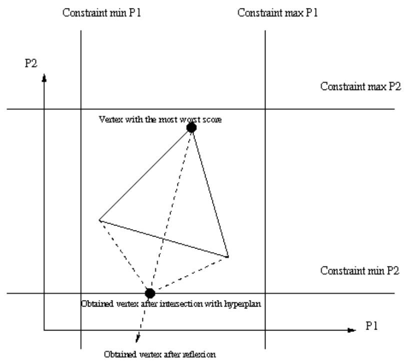
Figure 3 : Examples of biomechanical constraints.

domain is convex, any contraction of a simplex also satisfies the constraints. Consequently, constraints only need to be checked after symmetries or expansions. If a vertex happens to be moved outside the constraints domain during one of these transformations, then it is replaced with a new vertex at the intersection of the violated constraint hyperplane and the line in the move direction starting from the initial vertex (figure 4).

The initial simplex is bound in a similar way. Note that, to increase the chances of reaching the global optimum, the initial simplex should be as large as possible inside the constraints domain. One of the vertex is an initial guess of the vector of parameters possibly resulting from temporal tracking in the image sequence. A simple prediction could be the registration results at the previous image. The initial simplex is then built by adding vertices along the axes of the parameter space, as far as possible from the initial guess with respect to the constraints. Both directions of the parameter axes are tried.



(a) The constrained simplex



(b) Constrained symmetry

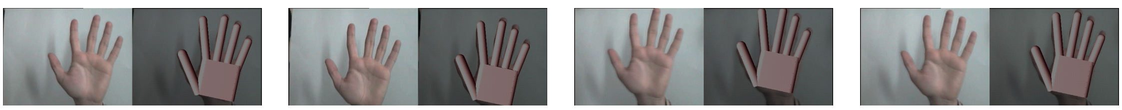
Figure 4 : An example of constraining the simplex transformations.

9 Experimentations and results

This approach has been tested on image sequences showing translations of the hand or abduction/adduction and flexion/extension of fingers. Results are shown in figures 5 through 9. Most of them demonstrate correct registration of the model.
However the following restrictions have been considered. First, our model does not include forearm to avoid bias on registration. Hence if we use non-overlapping surface, we interactively remove forearm from the image. Notice that wrist localisation can be automated [17]. Second, as global rotations, scale factor and interphalangeal flexions may cause ambiguity, they were not considered in this experimentation.
To remedy to this shortage, we have developed a hierarchical technique that consists in first registering only the global position, then successively register scale factor, rotation DOF, abduction and MCP flexion. These 16 DOF lead to a less accurate registration. We can enhance accuracy by including information on optical flow in the registration procedure to choose the order of the hierarchical registration.
In case of flexion, it may happen that the third and second phalanx of a finger are either occluded by the rest of the hand, or appear inside the silhouette. Because the cost functions described so far do not provide any information on these parts of the hand, neither optimisation method can register them correctly. We are currently considering using edges inside the silhouette to cope with fingers bent in front of the palm.
The sequences are not processed in real time but optimisation of source code and automatic choice of thresholds to stop iterative optimisation procedure will decrease the computation time.

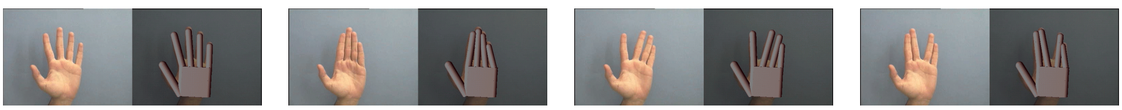
Conclusion and perspectives

In this paper, we have compared several registration evaluation functions and optimisation method for 3D hand model registration. The non overlapping surface with downhill simplex performs better than any other combinations between optimisation methods and cost functions. We integrated biomechanical constraints with downhill simplex minimisation algorithm. This considerably reduces the search space and excludes unrealistic postures. We have demonstrated the validity of our approach through tests conducted on natural sequences, and discussed its limits. A persistent shortage of our approach is its inability to process DOF leading into ambiguity. Solving this problem will be the aim of our future work.



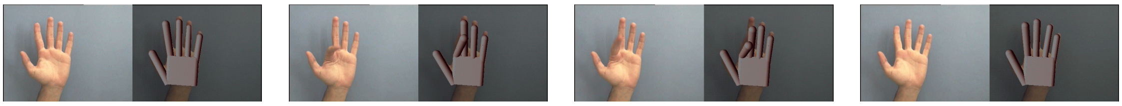
(a) image 0 (b) image 7 (c) image 38 (d) image 47

Figure 5 : Hand translation.



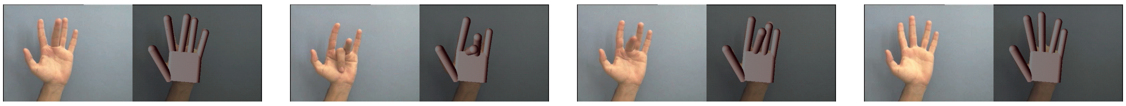
(a) image 0 (b) image 18 (c) image 55 (d) image 67

Figure 6 : Abduction/adduction fingers.



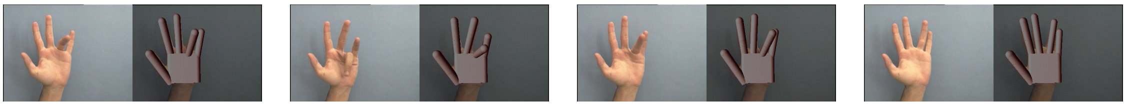
(a) image 0 (b) image 13 (c) image 34 (d) image 44

Figure 7 : Flexion index.



(a) image 10 (b) image 27 (c) image 35 (d) image 41

Figure 8 : Flexion middle.



(a) image 12 (b) image 26 (c) image 44 (d) image 50

Figure 9 : Flexion ring.

References

[1] A. Braffort, *Reconnaissance et Compréhension de gestes, application à la langue des signes*, thèse de l'université de Paris 11, spécialité informatique, juin 1996.

[2] J. Martin and L. Crowley, "An Appearance-Based Approach to Gesture Recognition", *proc. of the 9th Conf. on Image Analysis and Processing*, Italy, 1997.

[3] J. M. Regh and T. Kanade, "Visual Tracking of Height DOF Articulated Structures : an Application to Human Hand Tracking", *Proc. 3rd ECCV*, Vol. 2, 1994, pp. 37-46.

[4] H. Ouhaddi , P. Horain, "Conception et ajustement d'un modèle 3D articulé de la main", *Actes 6~èmes Journées de Travail du GT Réalité Virtuelle*, France, Mars 1998.

[5] M. Mochimaru and Y. Yamazaki, "The three-dimensional measurement of unconstrained motion using a model-matching method", *Ergonomics*, Vol. 37, No 3, 1994, pp. 493-510.

[6] J. J. Kuch and T. S. Huang, "Vision Based Modeling and Tracking for Virtual Teleconferencing and Telecollaboration", *Proc. of 5th Int. Conf. On Computer Vision*, June 1995. pp. 666-671.

[7] J. Ohya and F. Kishino, "Human Posture Estimation from Multiple Images Using Genetic Algorithm", *proc. of the 12th IAPR Int. Conf. on Pattern Recognition*, Vol I, pp. 750-753.

[8] D.M. Gavrila and L. S. Davis, "3D model based tracking of humans in actions : a multi-view approach", *proc. of IEEE Computer Vision and Pattern Recognition*, 1996.

[9] Shimada, Y. Shirai, Y. Kuno and J. Miura, "Hand Gesture Estimation and Model Refinement using Monocular Camera-Ambiguity Limitations by Inequality Constraints", *proc. of FG'98*, pp. 268-273, 1998.

[10] G. Borgefors, "Distance transformation in digital images", *Computer Vision, Graphics and Image processing*, vol. 34, 1986, pp. 344-371.

[11] H. Ouhaddi, P. Horain, K. Mikolajczyk, "Modélisation et suivi de la main", *Actes 4èmes Journées d'études et d'échanges "Compression et REprésentation des Signaux Audiovisuels" (CORESA'98)*, France, June 1998, pp. 109-114.

[12] D. G. Lowe, "Fitting Parametrized Three-Dimensional Models To Images", *IEEE Transactions on Pattern and Machine Intelligence*, Vol. 13, no 5, May 1991, pp. 441-450.

[13] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization", *Computer Journal*, Vol. 7, 1965, pp. 308-313.

[14] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, N*umerical Recipes in C*, Cambridge University Press, 1992.

[15] K. I. M. McKinnon, C*onvergence of the Nelder-Mead Simplex Method to a non-stationary point*, technical report MS96-000, Department of Mathematics and Statistics, University of Edinburgh, May 1996.

[16] S. Wachter and H. H. Nagel, "Tracking of Persons in Monocular Images Sequences", *proc. IEEE Non Rigid and Articulated Motion Workshop*, pp. 2-9, 1997.

[17] B. Deimel and S. Schroter, "Improving Hand-Gesture Recognition via Video Based Methods for the Separation of the forearm from the Human Hand", *Gesture Workshop'99*, France, March 1999.