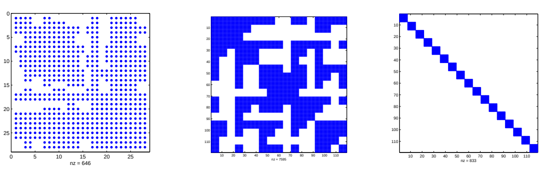
Model-Based Hand Tracking with Texture, Shading and Self-occlusions: Supplementary material

Martin de La Gorce[1] - Nikos Paragios[1,2] - David J. Fleet[3]

1. Laboratoire MAS, Ecole Centrale de Paris
2. Equipe GALEN, INRIA Saclay - Ile-de-France, Orsay, France
3. Department of Computer Science, University of Toronto, Canada

1. Blockwise BFGS update



(a) (b) (c)

Figure 1. Sparsity structure: (a)$\theta \in \mathbb{R}^{28}$ (b) $\theta \in \mathbb{R}^{119}$ (c) approximated structure

Quasi-Newton methods are iterative minimization methods that maintain, through iteration, a convex quadratic function that locally approximates the function to be minimized. In our case, each iteration would consist in

1. Estimating the minimum $\hat{\theta}$ of the quadratic approximation under the set of linear constraints corresponding to joint limits. This is a convex quadratic programming problem that can be solve using standard methods.

2. Estimating θ_t by performing a line search on the half-line $\{1-\alpha)\theta_{t-1} + \alpha\hat{\theta}|\theta > 0\}$.

3. Updating the quadratic model by using the gradient at θ_t and by updating the Hessian with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula.

We have to carefully initialize the approximated Hessian to obtain good convergence rate during the first iterations. In order to do so, we used a scaled version of the matrix ${J_\theta^{\bar{V}}}^t J_\theta^{\bar{V}}$ where $J_\theta^{\bar{V}}$ is the Jacobian of projected vertices w.r.t θ. Such an initialization favors the displacements in the depth direction for which the gradient is small due to little data support.

Because the contributions to the overall cost of two well-separated fingers are independent, the true Hessian will not be fully populated but exhibit blocks of zeros (fig.1.a). The sparsity of the Hessian is accentuated if, instead of using joints angles, we parameterize individually the pose of each of the 17 bones using a 7D vector composed of a quaternion and a translation vector such that $\theta \in \mathbb{R}^{119}$ (fig.1.b). Non zeros entries of the 119 by 119 Hessian appears on 7 by 7 blocks. Each block that is not on the diagonal corresponds to a pair of hand parts that either occlude each other or share some facets in their influence area when the pose space deformation method is used. Using quaternions for the hand pose parameterization would facilitate the use of finger independencies in the minimization process. Unfortunately this would require additional non-linear equality constraints to enforce validity of relative poses between linked bones. Such non linear equality constraints are difficult to handle in a continuous optimization framework and are likely to decrease the convergence rate of the method. In order to exploit the sparsity of the Hessian when quaternion are used for the hand pose parameterization without the need to add non-linear constraints, we first decompose the function $E(\theta)$ into$E(\theta) = E_q(Q(\theta))$, where Q maps the joints-angles pose representation to the quaternion representation. The Hessian $\frac{\partial^2 E_q}{\partial^2 \theta}$ is then approximated by $(\frac{\partial Q}{\partial \theta})^t H_q (\frac{\partial Q}{\partial \theta})$ with $H_q = (\frac{\partial^2 E_q}{\partial^2 Q})$. At each step, we update an approximation of the Hessian H_q with an adapted BFGS update. We approximate the structure of H_q by assuming a complete independence between part of the hand. This result in a block-diagonal structure (fig.1.c) where non zeros entries are restricted to the 7 by 7 blocks along the diagonal. The standard BFGS update does not exploit the sparsity structure of the approximated Hessian and would populate the entire matrix with non-zeros values. Using the BFGS formula, we do not update the whole matrix H_q, but we update independently each non-zeros 7 by 7 blocks along the diagonal of the matrix. About 7 gradient evaluations are then necessary to obtain a reasonable local approximation of the Hessian when the standard BFGS method would require about 28 evaluations. This has a direct impact on the convergence rate of the minimization. The method induce more zeros values than in the true Hessian but still lead significant improvement over the standard BFGS update. As we keep performing increments on the θ vector during the optimization, we do not need to add non-linear constraints to enforce validity of relative poses between linked bones that would be necessary if increments where done in the quaternion representation space. The improvement in the minimization process, in term of number of call to the objective function, is illustrated in fig where we estimated the pose for a single frame.

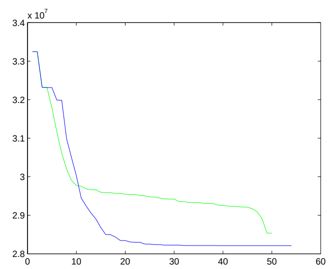


Figure 2. Functional decrease w.r.t number of functional evaluations. blue : adapted BFGS. green : normal BFGS