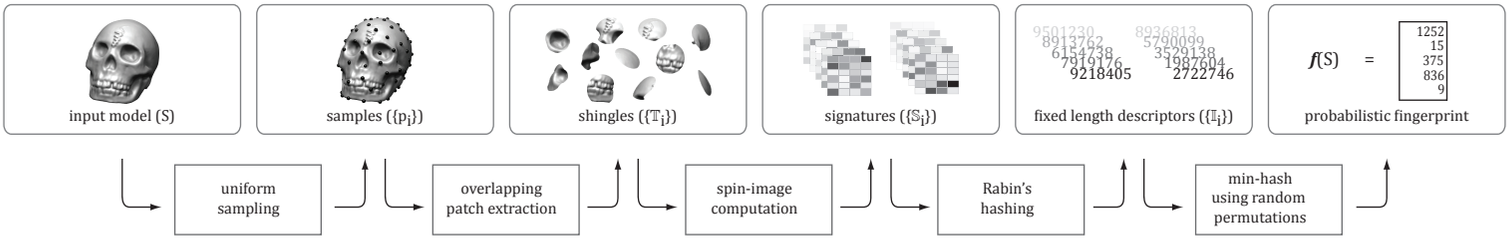
Probabilistic Fingerprints for Partial Shape Similarity

Niloy J. Mitra† Leonidas J. Guibas† Joachim Giesen‡ Mark Pauly§

†Stanford University ‡Max Planck Institut für Informatik §ETH Zurich



Shape Similarity

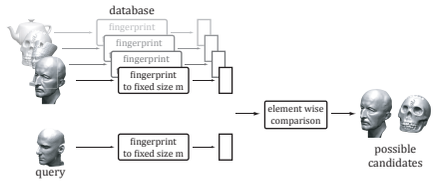
Given two shapes S_1 and S_2, their *distance* $D(S_1,S_2)$ satisfies the following:

(Identity)	$D(S_1,S_1) = 0$
(Symmetry)	$D(S_1,S_2) = D(S_2,S_1)$
(Rigid transform invariance)	$D(S_1,S_2) = D(\alpha(S_1),S_2)$
(Partiality)	$S_1 \subseteq S_2 \Rightarrow D(S_1,S_2) = 0$

Compute a *compact fingerprint* $\boldsymbol{f}$ such that:

$\boldsymbol{f}(S_1) \neq \boldsymbol{f}(S_2) \Rightarrow S_1$ and S_2 are dissimilar
$\boldsymbol{f}(S_1) = \boldsymbol{f}(S_2) \Rightarrow S_1$ and S_2 are similar
} with high probability
$\boldsymbol{f}$ is compact $\rightarrow |\boldsymbol{f}(S)| \ll |S|$
$\boldsymbol{f}$ is efficiently computable

Motivation



shape analysis → scan alignment, shape retrieval, partial matching, shape clustering, recognition } are two shapes similar?

Shingles

partial shape matching ⇒
comparing two *unordered* high dimensional point sets (spatial realtions lost) using *sketches*

robustness to mesh tesselation ensured using dense sampling $\delta \ll \rho$

Resemblance

DEFINITION

resemblance computed from signatures

$$r(S_1,S_2) = \frac{|\mathbb{S}_1 \cap \mathbb{S}_2|}{|\mathbb{S}_1 \cup \mathbb{S}_2|} = \frac{\sum_{\sigma \in \mathbb{S}_1 \cap \mathbb{S}_2} \min(m_1(\sigma), m_2(\sigma))}{\sum_{\sigma \in \mathbb{S}_1 \cup \mathbb{S}_2} \max(m_1(\sigma), m_2(\sigma))}$$

APPROXIMATION

min-hash [Broder '00]

$\pi_1, \ldots, \pi_m$: m random permutations on $\mathbb{U}$, the universe of all possible symbols

probabilistic fingerprint

$\boldsymbol{f}(S) = (\min\{\pi_1(\mathbb{I})\}, \ldots, \min\{\pi_m(\mathbb{I})\})$

resemblance *estimate* using fingerprints

$$r'(S_1,S_2) = \sum_j \min(m_{1j}, m_{2j})\ \chi(\boldsymbol{f}(S_1)_j = \boldsymbol{f}(S_2)_j) \ /\ \sum_j D_j$$

$D_j = \max(m_{1j}, m_{2j})\ \chi(\boldsymbol{f}(S_1)_j = \boldsymbol{f}(S_2)_j) + m_{1j}\chi(\boldsymbol{f}(S_1)_j < \boldsymbol{f}(S_2)_j) + m_{2j}\chi(\boldsymbol{f}(S_1)_j > \boldsymbol{f}(S_2)_j)$

Estimation Error

Pipeline Stage	Source of Error	Analyzed using
Rabin's hashing	many-to-one mapping	counting argument
min-hashing	random sampling	Chernoff bound

For sets, in order to have probabilistic error bound

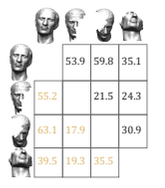
$$\Pr[(1-\delta)r(S_1,S_2) \leq r(S_1,S_2) \leq \lambda(1+\delta)(1+\sqrt{\varepsilon})r(S_1,S_2)] \geq 1-\eta$$

we have to choose,

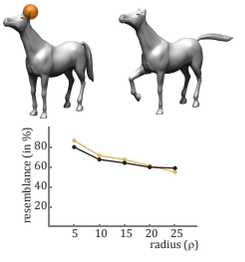
$$m \geq 4\ln(2/\eta)/(\delta^2 r(S_1,S_2))$$

For typical values of parameters, m≈1000 suffice.

Evaluation



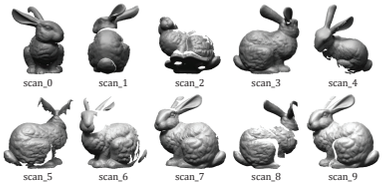
Comparison with ground truth : cumulative error due to Rabin's hashing and min-hashing



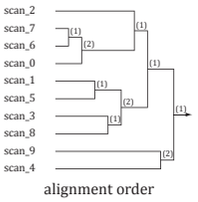
Effect of shingle size on resemblance

Applications

MULTIPLE SCAN ALIGNMENT



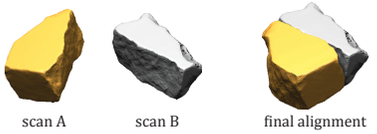
multiple scans in arbitrary initial positions



alignment order

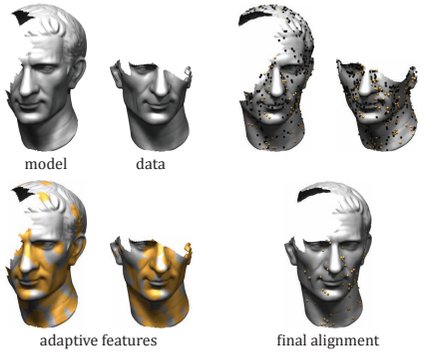
final alignment

COMPLEMENTARY SHAPE MATCHING



scan A scan B final alignment

PARTIAL SCAN ALIGNMENT



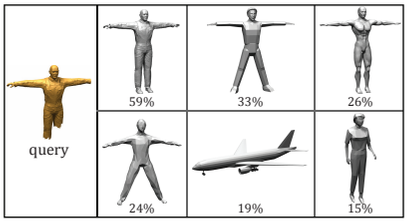
model data adaptive features final alignment

PERFORMANCE (fingerprint size 10kB)

model	# vertices	pre-proc. (in secs)	query time (in msec)
skull	54k	13	15
Caesar	110k	18	15
bunny	1,210k	16	15
horse	8k	7	15
database	(1,814 models) average 7k	average 5	average < ½ sec

SHAPE SPACE

DATABASE RETRIEVAL



For more information:
http://graphics.stanford.edu/~niloy/research

Symposium on Geometry Processing 2006