Supplemental materials for PoseShop

1 Verb and verb-object phrase list

akimbo	applaud	archery	argue	arrest	ask	athletics
attack	bake	balance	ballet	bathe	begging	blow
boast	body build	bounce	bow	box	brake	bump
bungee jump	bury	call	camp	carry	carry bags	carry books
carry box	carry papers	cartwheel	carve	chase	cheer	chew
choke	chop	circle kick	clap	climb	clip	comb
command	communicate	compete	complain	conduct	confess	cook
crash	crawl	creep	crouch	crush	cry	cut vegetables
curl	damage	dance	decorate	defense	destroy	dig
direct	disapprove	discover	dive	drag	drain	draw pictures
dress	drink	drink coffee	drive	drop	drown	drum
eat	eat apple	eat beef	eat cake	eat noodles	eat watermelon	entertain
excite	fall	feed bird	feed cat	feed dog	fence	fetch
fetch flower	fetch water	film	fire	fish	fling arms	fly
flying kick	fold arms	go downstairs	go upstairs	grab	grin	grip
gymnastics	haircut	handshake	hang	heal	high jump	high kick
hike	hip-hop	hit	hold book	hold box	hold cat	hold dog
hold fish	hold flower	hold out hand	hold water	hop	howl	hurdle
hug	inject	jog	judo	juggle	jump	kick
kickboxing	kick ball	kill	kneel	knife	knit	knock
kungfu	land	laugh	launch	lead	leap	learn
lick	lie	lie down	listen	lock	long jump	look
look up	marathon	march	match	mine	morning exercise	mourn
move	murder	nail	nest	nod	observe	on tiptoe
order	pack	paint	paper airplane	parachute	parallel bars	paste
pat	peck	peep	phone	pick	pick apple	pick up
pinch	plant	plant tree	play	play baseball	play basketball	play cards
play computer	play cricket	play football	play golf	play guitar	play handball	play hockey
play piano	play rugby	play soccer	play table-tennis	play tennis	play violin	play volleyball
point	pour	pray	preach	prick	print	propose marriage
pull	pull dog	pump	punch	push	push-ups	race
raise	raise arm	raise hand	raise leg	reading	refuse	regret
reject	relax	repair	report	reproduce	ride bicycle	ride horse
ride motorcycle	rinse	rob	rock	rock'n'roll	rock climbing	roll
rotate	rub	run	rush	sail	saw	scare
scold	scream	scribble	scrub	search	serve	shave
shiver	shock	shoot	shoot rifle	shop	shout	shrug
sigh	sign	sing	sip	sit	sit-ups	skate
skateboard	ski	skip	slap	sleep	slide	slip
smell	smile	smoke	snatch	sneak	sneeze	sniff
snore	soak	spear	squat	squeal	stand	stare
step over	stomp	stoop	stretch	stretch hand	stretch out	suck
surf	surrender	surround	swim	swing	sword	taekwondo
take a knee	take medicine	take photo	tango	talk	taste	throw
throw ball	throw darks	throw discus	throw javelin	throw stone	thumb down	thumb up
tiptoe	toothbrush	touch	tour	tread	trip	tumble
twist	undress	unfasten	unlock	unpack	walk	wander
wash cloth	wash face	watch	watch TV	wave hand	weightlift	whisper
wipe	work	wrestle	yawn			

Table 1: Verb and verb-object phrase list, presented in alphabet order.

2 Original image components

Shown are the original image components used to generate the comics included in the paper. Human characters are extracted from them and composed with other scene items and backgrounds.

Figure 1: Original image components used in the paper.

3 Sample images in our database

Figure 2: Sample images associated with ‘girl run’, ‘woman jump’, ‘man play football’ and ‘boy skating’. In each sub-figure, the four rows show images with one of the four cloth attributes, i.e. ‘long sleeves + shorts’, ‘T-shirts + long pants’, ‘long sleeves + long pants’ and ‘T-shirts + shorts’. The black-and-white images in the first row are representative poses. The segmentation is overlaid in the image. Images with incorrect content (without a character of the right action) are framed in yellow.

4 Sample results of skin detection

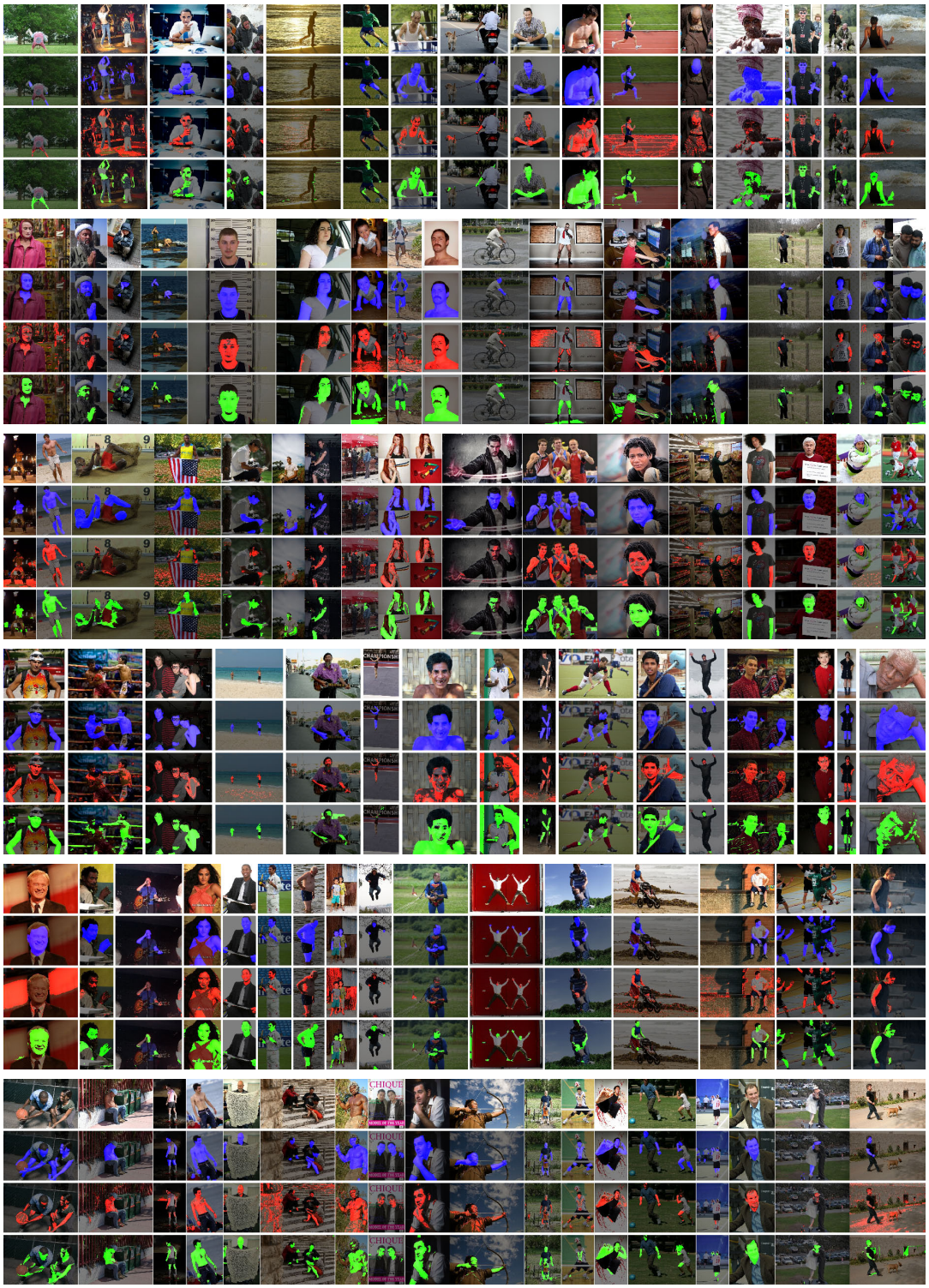

Figure 3: In each sub-figure, there are four rows corresponding to the input images, skin pixels (blue) manually marked by user, detected skin pixels (red) according to [Jones and Rehg 2002], detected skin pixels (green) according to our method. The false skin and non-skin ratio reported in our paper is all the fraction of the amount of *skin* pixels. If we treat the non-skin ratio to be fraction of the amount of non-skin pixels, it will be 4.9% and 2.9% for [Jones and Rehg 2002] and our method. That shows non-skin pixel amount is a lot larger than skin pixel amount. Note in [Jones and Rehg 2002], the accuracy of skin detection is reported to be 88%, but at this threshold, the false non-skin ratio (to all the non-skin pixels) will be about 11%. While this is desirable for other applications, our segmentation cannot tolerate such a false negative rate.

5 Segmentation Evaluation

src										
our										
gtr										
TP(%)	97.71	93.55	95.31	98.88	97.77	91.47	95.33	93.60	60.30	88.56
FP(%)	0.15	0.051	0.075	0.064	0.048	0.077	0.047	0.15	0.079	0.045
src										
our										
gtr										
TP(%)	80.13	98.61	99.08	99.56	97.97	84.75	98.81	69.56	86.50	85.10
FP(%)	0.94	0.046	0.66	0.12	0.70	0.76	0.14	0.13	4.26	0.12
src										
our										
gtr										
TP(%)	99.98	98.62	91.67	98.96	98.75	91.86	98.78	99.19	89.33	97.06
FP(%)	0.13	0.35	0.65	0.43	0.042	3.28	2.45	0.25	0.13	1.10
src										
our										
gtr										
TP(%)	82.28	93.63	74.78	92.65	94.52	99.76	98.11	84.05	88.19	92.66
FP(%)	0.51	0.13	0.041	0.92	1.48	0.91	0.26	0.14	2.23	0.28
src										
our										
gtr										
TP(%)	83.70	97.61	92.43	86.14	99.33	91.39	96.07	61.64	55.18	98.40
FP(%)	0.027	0.23	2.67	3.06	2.42	0.029	3.45	6.51	0.029	2.45

src										
our										
gtr										
TP(%)	100.00	99.31	99.12	91.42	99.61	99.42	93.40	89.23	93.44	99.64
FP(%)	0.00	0.064	0.83	0.072	3.19	0.10	1.04	1.92	0.61	0.52
src										
our										
gtr										
TP(%)	93.79	97.22	96.26	93.22	91.42	98.42	98.75	96.99	97.68	95.55
FP(%)	0.26	0.098	0.097	1.42	0.72	0.63	0.47	0.24	0.18	0.028
src										
our										
gtr										
TP(%)	94.95	82.11	99.50	79.06	100.0	99.47	96.74	89.48	94.90	97.84
FP(%)	1.04	3.19	0.43	2.11	0.00	0.48	5.40	0.21	4.74	0.12
src										
our										
gtr										
TP(%)	85.49	90.82	95.00	82.97	99.07	91.71	92.43	57.59	69.98	99.50
FP(%)	1.95	4.95	0.13	2.64	0.58	6.66	0.51	0.079	2.72	0.52
src										
our										
gtr										
TP(%)	93.15	82.43	98.49	95.53	86.05	77.50	95.42	93.26	73.15	97.37
FP(%)	1.86	0.068	3.04	0.73	0.21	0.076	0.25	4.77	0.45	1.46

src										
our										
gtr										
TP(%)	98.86	84.69	90.06	90.41	95.20	91.37	91.33	93.57	98.41	87.55
FP(%)	0.035	0.23	0.46	5.78	0.12	0.094	0.17	0.15	0.11	0.81
src										
our										
gtr										
TP(%)	94.28	92.44	91.16	96.90	96.57	99.91	93.85	97.10	93.43	95.70
FP(%)	0.32	0.43	0.91	0.42	0.32	1.94	0.018	0.65	0.023	0.19
src										
our										
gtr										
TP(%)	95.39	83.65	62.34	79.90	81.39	74.13	94.77	73.16	98.33	92.19
FP(%)	0.29	0.24	0.009	4.45	0.60	3.80	1.06	0.016	0.72	1.60
src										
our										
gtr										
TP(%)	91.48	99.11	92.33	80.27	91.92	96.07	97.85	95.92	72.72	74.84
FP(%)	0.57	0.10	0.077	0.073	4.57	0.080	0.34	0.17	0.006	0.72
src										
our										
gtr										
TP(%)	99.74	83.66	87.49	95.45	89.09	90.24	85.87	71.65	96.30	63.16
FP(%)	0.23	1.39	0.077	1.04	2.75	0.37	0.068	0.82	0.43	0.12

src													
our													
gtr													
TP(%)	90.70	87.06	78.47	89.81	86.92	83.23	55.84	99.91	99.62	67.33	99.63	84.23	98.58
FP(%)	0.54	0.041	2.12	0.010	0.36	0.057	0.064	4.60	0.52	0.20	0.19	0.088	0.23
src													
our													
gtr													
TP(%)	55.44	87.25	86.82	83.17	92.19	96.21	94.08	82.70	84.45	98.63	99.64	95.45	92.69
FP(%)	2.39	0.06	0.94	0.17	0.14	0.047	0.87	3.98	2.42	0.45	0.078	0.44	0.21
src													
our													
gtr													
TP(%)	93.26	82.94	81.66	85.40	85.63	67.11	94.26	32.64	97.44	49.66	98.02	96.16	
FP(%)	4.83	0.056	0.077	0.39	2.74	0.030	6.08	0.086	1.00	0.40	0.18	2.73	
src													
our													
gtr													
TP(%)	99.46	94.75	95.27	75.87	76.05	44.62	68.40	90.88	90.18	92.22	94.23	98.77	
FP(%)	10.37	3.70	3.70	1.05	0.048	0.17	0.20	0.13	0.008	0.70	0.081	0.18	

Figure 4: Per-pixel segmentation quality. These 200 images are uniformly sampled from our database associated with "boy skating", "girl run", "man play football" and "woman jump". "src", "our" and "gtr" are original online images, segmentation masks in our database and manually labeled ground truth segmentation masks respectively. "TP" and "FP" are true positive rates and false positive rates respectively. The average true positive rates (the ratio of detected human pixels) are 89.2%, 91.5%, 88.5% and 83.9%. The average false positive rates (the ratio of non-human pixels labeled as human) are 0.9%, 1.2%, 0.8% and 1.2%.

6 More Results and Comparisons

6.1 More Results

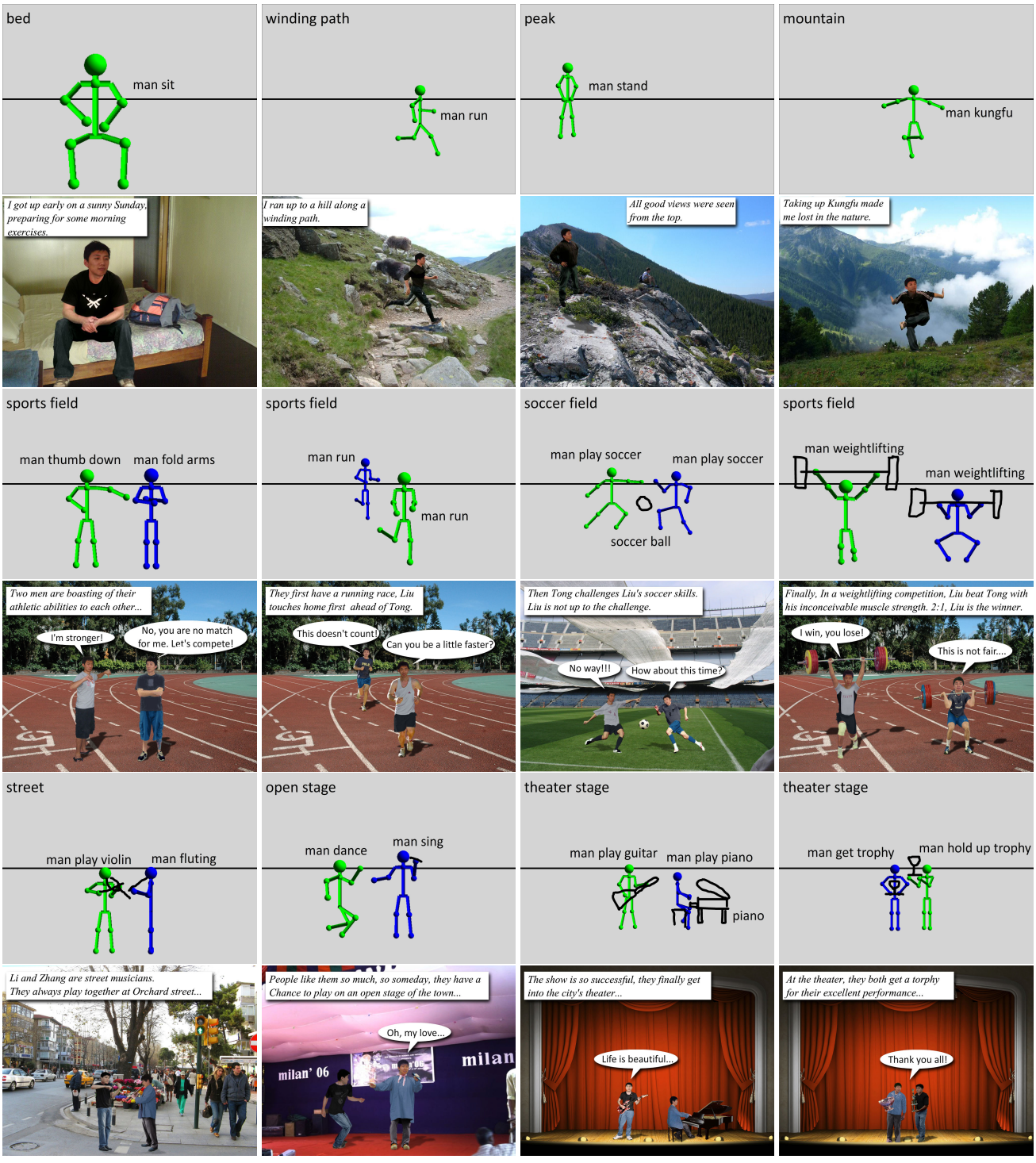

Figure 5: More comic-strips generated by a novice user to our system after 30 minutes training.

6.2 Compositions from a PhotoShop expert

Figure 6: Similar comic-strip compositions generated by a PhotoShop expert user who has more than 8,000 hours PhotoShop experience.

6.3 Comparisons

	morning	sports	band	beach	snow	kids	hunting
PS(min)	140+90	175+95	150+70	235+125	180+85	155+120	165+110
OUR(min)	3+12	5+15	5+13	3+11	2+7	2+6	4+10
PS(score)	3.2	3.8	3.4	4.2	4.4	4.4	3.8
OUR(score)	3.6	3.8	4.0	4.2	4.0	3.8	3.6

Table 2: Manual interaction time and composition quality comparisons. The interaction time of PhotoShop expert consists of two parts — one for finding image materials and the other for rotoscoping and composition. The interaction time of our system also consists of two parts — one for providing sketches and the other for interactive refinement. The bottom two rows are the average quality scores from 5 evaluators who are unaware of the generating process of these comic-strips. These subjective scores rang from 1 to 5. Scores: 5 = very good, photo-realistic; 4 = good, but with slightly noticeable artifacts; 3 = obvious artifacts, but still acceptable; 2 = very obvious artifacts, not quite acceptable; 1 = very bad, unacceptable.