

Our paper proposes a vision-based hand gesture recognition system with interactive training, aimed to achieve a user-independent application by on-line supervised training. Usual recognition systems involve a preliminary off-line training phase, separated from the recognition phase. If the system recognizes unknown (non-trainer) users the recognition rate of gesture classes could decrease. The recognition has to be suspended and all gestures need to be retrained with an improved training set, resulting in inconveniences. Our new approach introduces an on-line training method embedded into the recognition process, being interactively controlled by the user and adapting to his/her gestures. Our main goal is that any non-trainer users be able to use the system instantly and if the recognition accuracy decreases only the faulty detected gestures be retrained realizing fast adaptation. We implement the proposed system as a camera-projector system in which users can directly interact with the projected image by hand gestures, realizing an augmented reality tool in a multi-user environment. The emphasis is on the novel approach of dynamic and quick follow-up training capabilities instead of handling large pre-trained databases. We also conducted tests on several users in real environments for a practical application.

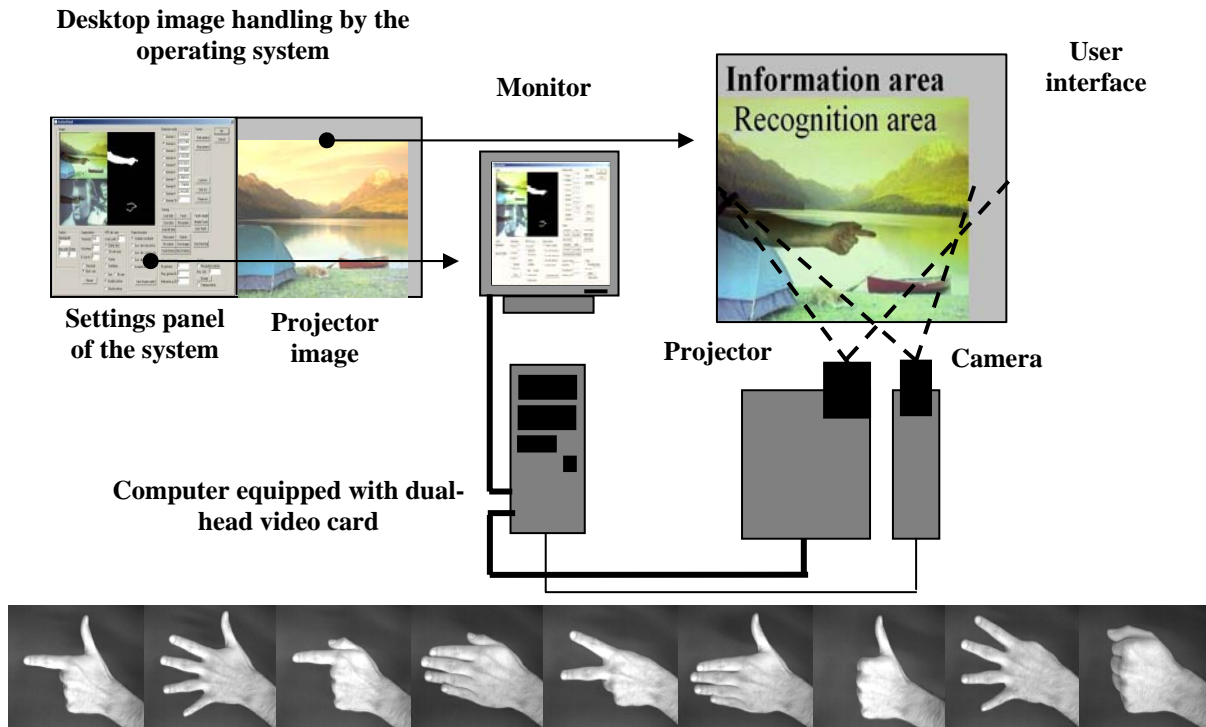


Fig. 1. System configuration and gesture vocabulary.

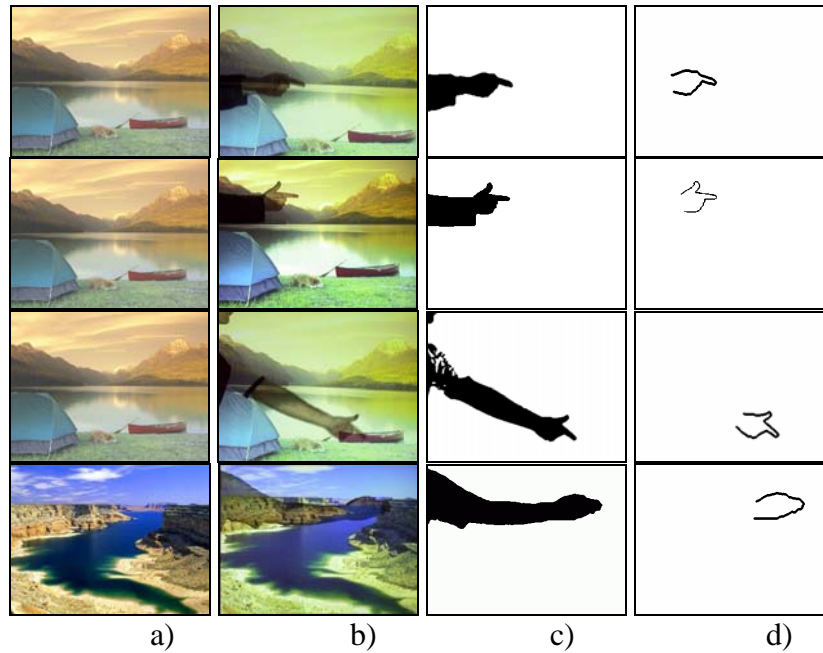


Fig. 2a-d. Results of the hand segmentation method wearing different types of clothes; a) projector image; b) camera image with arm; c) segmented arm image; d) extracted hand contour.

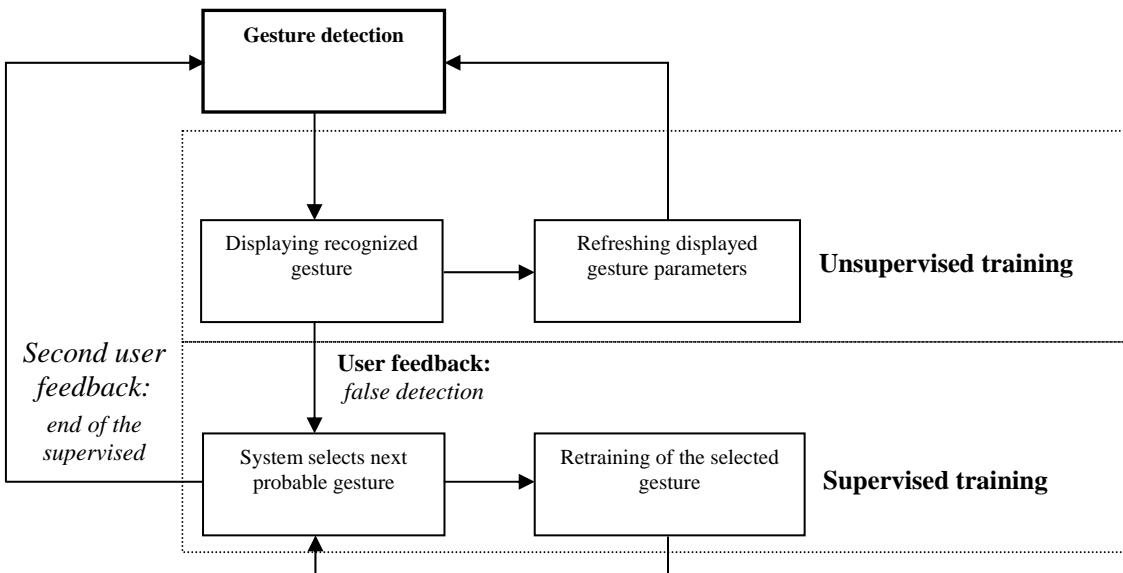


Fig. 3. Interactive training algorithm

Table 1 Test results with several trainer and tester users and improvements of the recognition by ML-based method; mean and standard deviation of the class recognition rates

Recognition results [%]											
Trainer users											
	<i>User A</i>		<i>User B</i>		<i>User C</i>		<i>User D</i>		<i>User E</i>		
Tester	Mean	S.d.	Mean	S.d.	Mean	S.d.	Mean	S.d.	Mean	S.d.	
<i>User A</i>	99.3	2	96.7	6.8	97.7	4.1	92.4	21.1	96.1	9.7	
ML-based	99.5	1.3	99.5	1.5	99.5	1.5	91.9	22.6	98.1	4.1	
<i>User B</i>	99	3.1	99.3	2.1	99.1	2.6	98.1	3.7	94	12.6	
ML-based	99.3	2	99.6	1.4	99.3	2.4	99	3	96.4	7.9	
<i>User C</i>	87.7	17.8	94.7	10.9	98.9	1.9	89.6	23.4	89.9	25.8	
ML-based	93	10.7	96.6	6.7	99.7	1	92.1	21.8	89.3	27.1	
<i>User D</i>	94.9	9.3	98.9	2.3	96.4	7.2	99.2	1.6	89.7	29.6	
ML-based	96.7	5.9	99.1	1.8	99.6	1.2	99.6	1.2	89.3	30.8	
<i>User E</i>	94.7	10.2	95.4	12.4	91.6	15.2	92.1	18.7	99.6	0.7	
ML-based	95.9	8	95.4	12.4	92.8	15.3	92.7	18.5	99.8	0.6	

Table 2 Recognition results in the interactive-based gesture recognition system tested by several users (“-” means that the recognition accuracy did not change)

User order	Recognition rates before the interactive training [%]		Recognition rates after the interactive gesture correction [%]		Number of supervised training cycles	Number of retrained gestures
	Mean	S.d.	Mean	S.d.		
	<i>User A</i>	99.3	2	-		
ML-based	99.5	1.3				
<i>User B</i>	99	3	-	-	0	0
ML-based	99.3	2				
<i>User C</i>	87.7	17.8	98	3.7	3	5
ML-based	93	10.7	99.6	1.2		
<i>User B</i>	98.9	12.8	98.9	3	1	2
ML-based	92.9	17.8	99.1	2.8		
<i>User D</i>	95.3	13	98.9	2	1	1
ML-based	96.1	10.6	99.6	0.9		
<i>User A</i>	98.1	2.9	-	-	0	0
ML-based	99.3	2				
<i>User C</i>	97.5	5.2	-	-	0	0
ML-based	98.4	4				
<i>User E</i>	93.1	11.8	99.5	1.1	3	4
ML-based	94.5	10.4	99.8	0.6		
<i>User B</i>	94.7	12.8	99.3	1.9	1	2
ML-based	96.4	7.9	99.4	1.7		