

# **Real-Time and Realistic 3D Facial Expression Cloning**

**Samsung Advanced Institute of Technology,**

**Multimedia Lab**

**Jung-Bae Kim**

September 10, 2010

## ■ 3D MR @ Home - Realistic 3D Mixed Reality



V-Sport

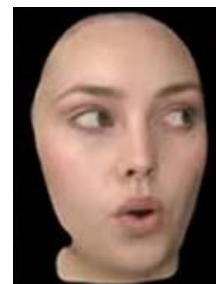
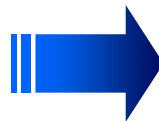


V-Entertainment



V-Commerce

## ■ Goal: To clone a user's natural expression to his avatar



# Problem Definition: Limitation of Previous Cloning Methods

- **Method 1: Highly Accurate Expression Cloning used in Movie CG**  
→ No real-time and cumbersome



Motion capture device of Vicon co.  
(Expensive 12 IR cameras are used. About 100 IR markers are attached on face. Several days are needed to clean up the internal noise.)

- **Method 2: Real-time Expression Cloning for Teleconference**  
→ Low accuracy

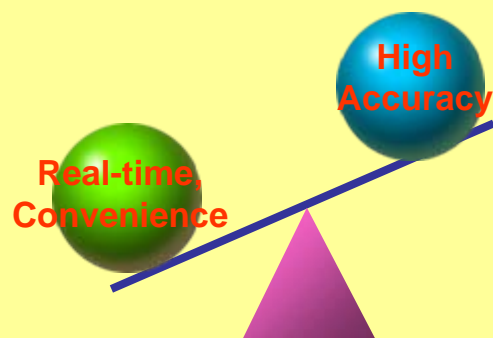


Teleconference S/W of Logitech (Frontal and symmetric expression)

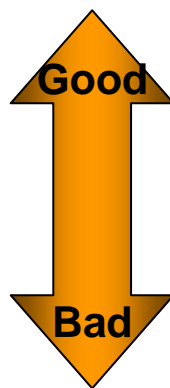
# Technical Contradiction



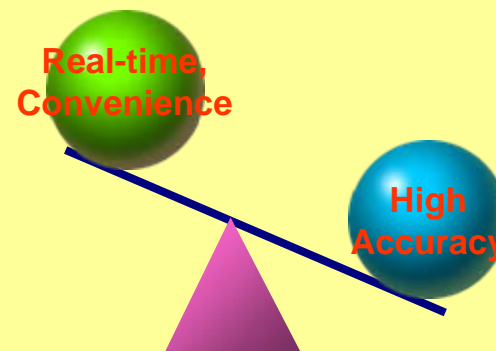
**TC 1 : Mocap-based method**



**(Method 1)**



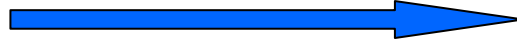
**TC 2 : Vision-based method**



**(Method 2)**

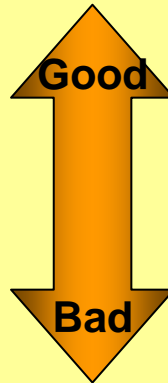


## Main IFR



## Sub IFR

IFR = Advantage of vision-based method  
+ Advantage of mocap-based method



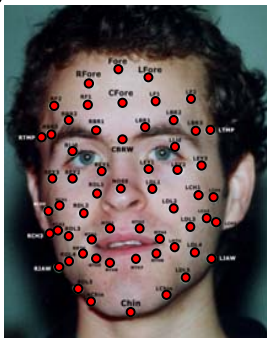
(New Method)

Sub-  
IFR1

<Face> itself does  
<put or remove markers on  
expression control points>

Sub-  
IFR2

<System> itself decides  
<the level of accuracy>



Markers  
on face

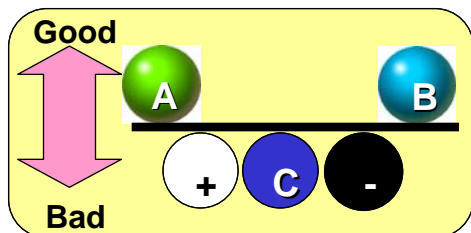
exist

- **High accuracy:** To be able to track variance among users, 3D head rotation, and expression change

not exist

- **Real-time:** Post-processing is not needed.
- **Convenience:** Only 1 camera is used. No need to attach cumbersome markers.

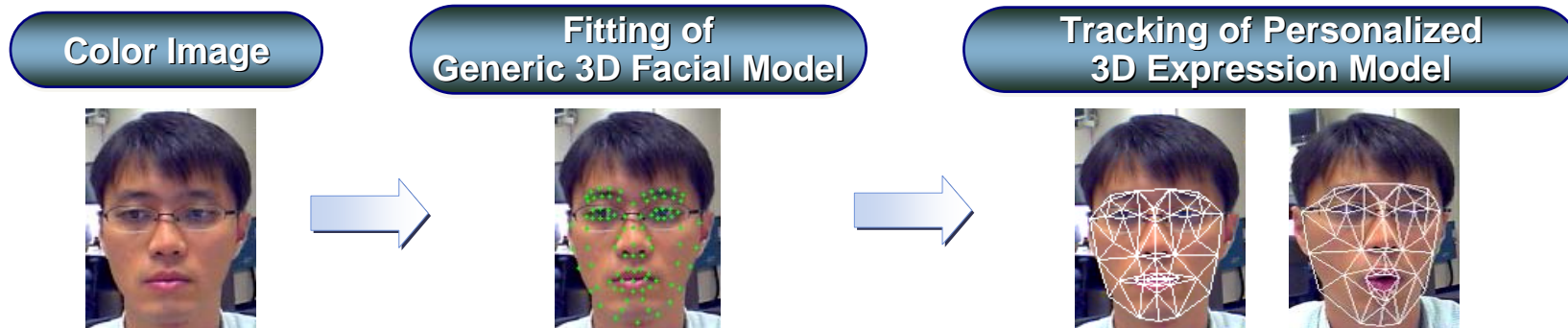
## ■ Physical Contradiction



For <accurate tracking>,  
<DoF on change> should be **<big>**,  
for <real-time processing>,  
<DoF on change> should be **<small>**.

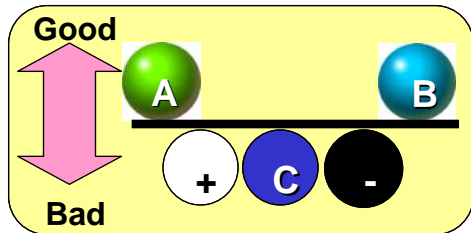
➔ DoF (Degree of Freedom) on Change: Inter-person change, intra-person change  
(head motion, expression)

## ■ Separation in Time



**In the initial step**, we consider only inter-person change,  
**after that**, we consider only intra-person change.

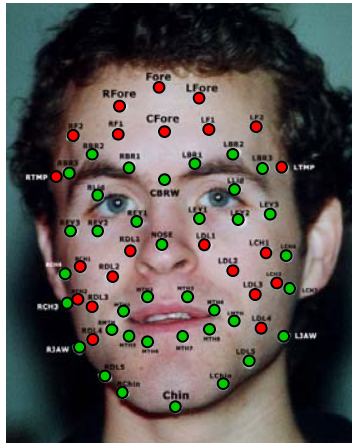
## ■ Physical Contradiction



For <real-time processing and convenience>,  
<markers on face> should **<not exist>**,  
For <high accuracy>,  
<markers on face> should **<exist>**.

## ■ Separation in Space (or Separation in Condition)

Expression  
control points:

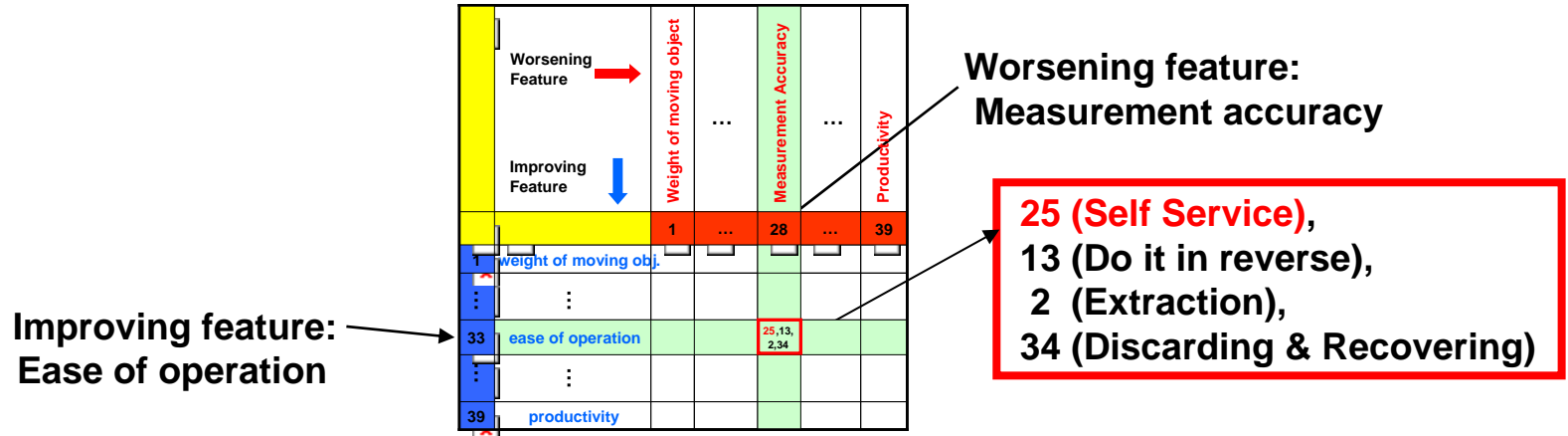


Track-able points by using vision method

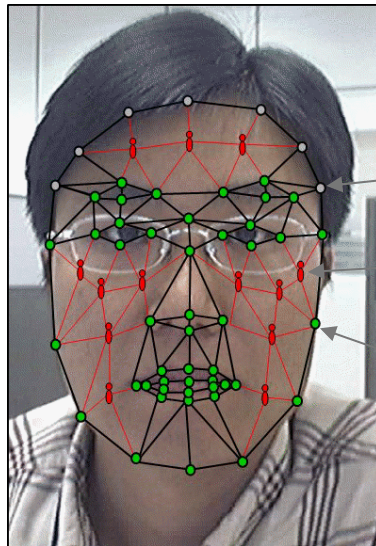
Track-able points by using Mocap method

**However, how Mocap data can be generated?**

## ■ Utilizing the Contradiction Matrix



## ■ Invention Principle 25 (Self Service), Smart Little Creature Model



Fixed points

Smart Little Creature

Tracked points  
by using vision-  
method

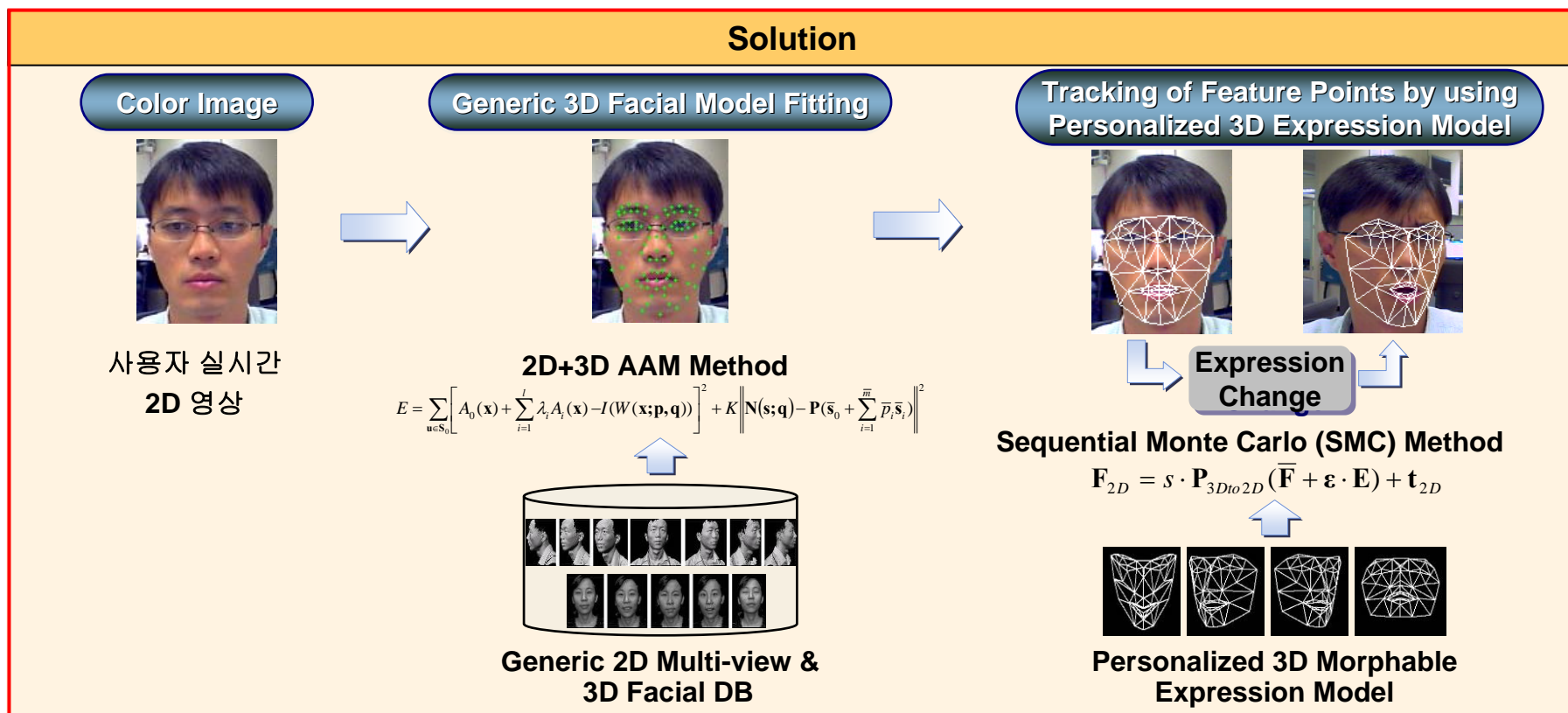
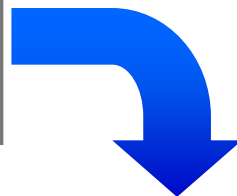
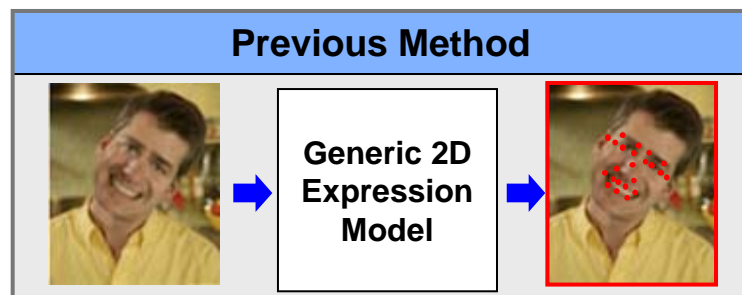
The system generates the motion of expression control points on forehead and cheek using Smart Little Creature (Muscle model).



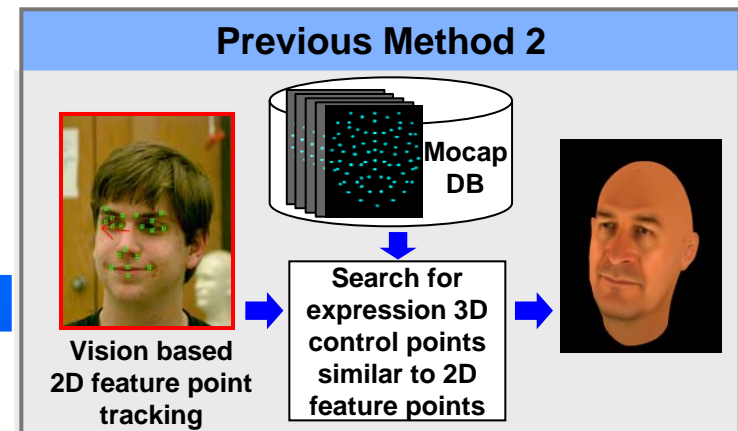
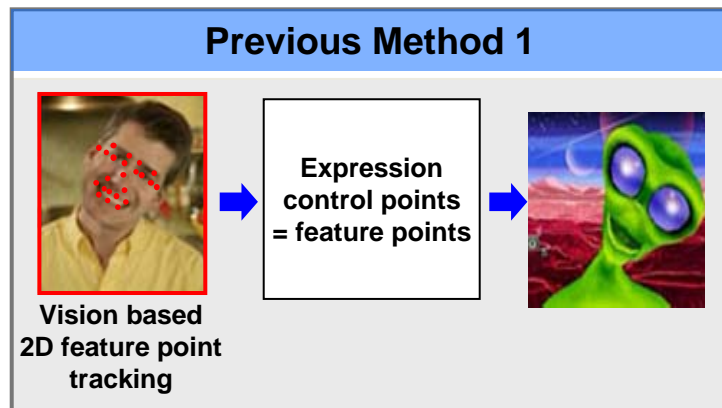
The muscle model is trained based on Mocap data.



# Solutions (1/2): Tracking of Feature Points by using Personalized 3D Expression Model

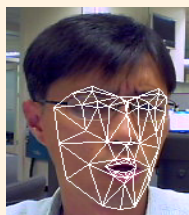


# Solutions (2/2): Generation of Expression Control Points based on Vision and Muscle Model



## Solution

Tracking of Feature Points by using Personalized 3D Expression Model

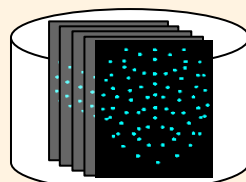


3D feature points

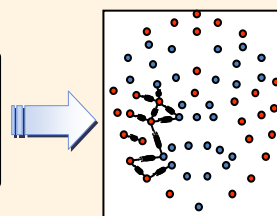
Generation of Expression Control Points based on Muscle Model

Expression Control Points = Feature Points + Muscle Model

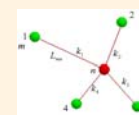
$$x'_n \approx \frac{\sum_m k_m \left( x'_m - \|L^0_{mn}\| \frac{L^{i-1}_{mn}}{\|L^{i-1}_{mn}\|} \right)}{\sum_m k_m}$$



Mocap DB



Muscle model based on stiffness

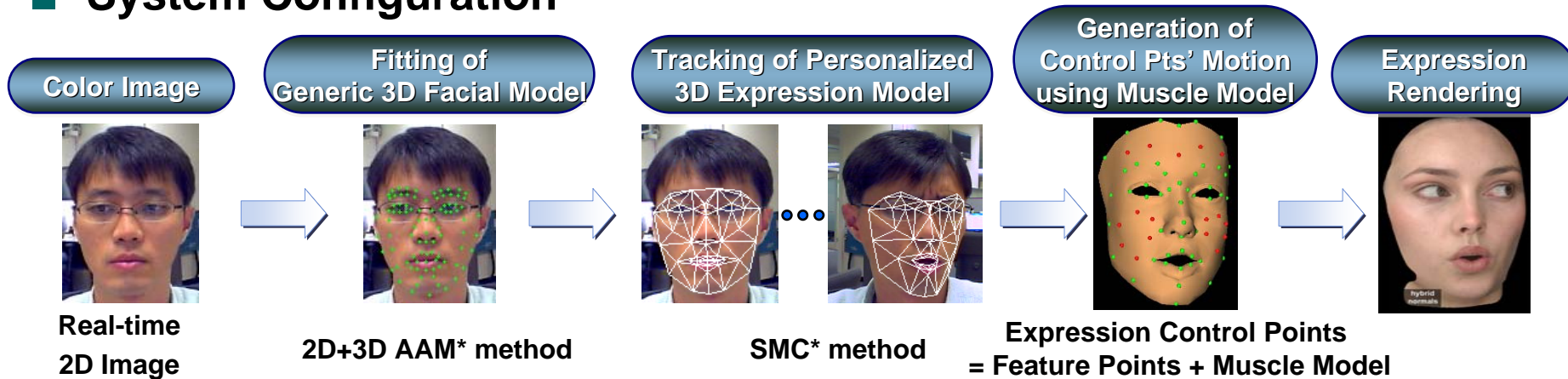


$$\sum_m k_m \left[ L'_{mn} - \|L^0_{mn}\| \frac{L'_{mn}}{\|L'_{mn}\|} \right] = 0$$

Expression Rendering



## System Configuration



## Performance

	Mocap-based Method	1 Camera Vision-based Method	Proposed Method
No. of Cameras	$\geq 7$	1	1
No. of attached markers on face	$\geq 70$	0	0
Processing speed	Offline	15fps	38.3fps
Track-able expressions	All expressions	Symmetric expression	All expression
Angle of head rotation	$-90^{\circ} \sim 90^{\circ}$ (3D Pose)	$-15^{\circ} \sim 15^{\circ}$ (2D Pose)	$-90^{\circ} \sim 90^{\circ}$ (3D Pose)
No. of control points	$\geq 60$	22	75

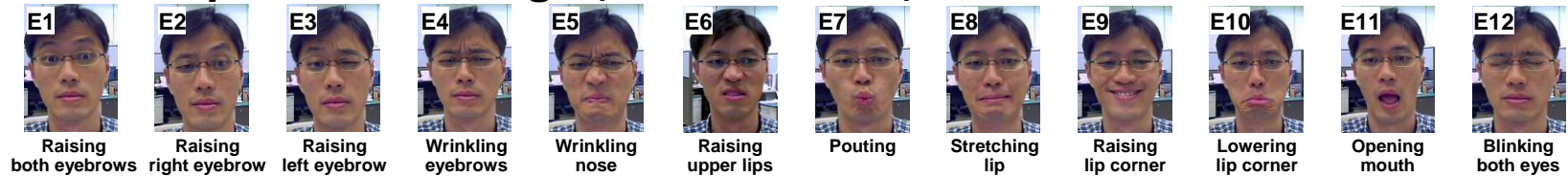
Convenience

Real-time

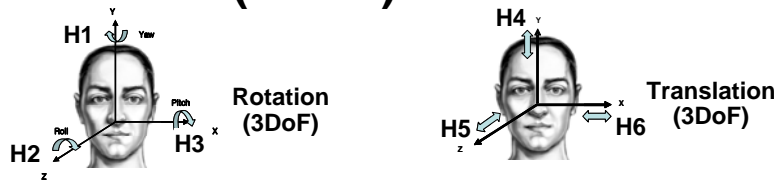
High accuracy

# Experimental Results

## ■ Various Expression Change (12 Action Units)



## ■ Head Rotation (6 DoF)



## ■ Recognition Ratio: 90.5% (10 persons, 504 expressions) @rotation< $\pm 90^\circ$ ; Processing Time 26.1ms @1024×768 image

