

# آشکار سازی محل ویژگی های صورت و بکارگیری آن در کد نمودن بر مبنای

## مدل دوبعدی

### فرزاد زرگری اصل

کارشناس ارشد گروه مهندسی برق - دانشکده فنی - دانشگاه تهران

### محمود کمره ای

دانشیار گروه مهندسی برق - دانشکده فنی - دانشگاه تهران

### محمدرضا الشریف - بهمن دولتشاهی

استادیاران گروه مهندسی برق - دانشکده فنی - دانشگاه تهران

(تاریخ دریافت ۷۸/۴/۲۰، تاریخ تصویب ۷۸/۹/۲۷)

## چکیده

روش های کلاسیک کد نمودن تصویر از خصوصیات آماری تصاویر مانند همبستگی زمانی و مکانی استفاده می نمایند. با استفاده از ویژگی های ساختاری تصاویر روش های جدیدی برای کد نمودن و فشرده سازی تصاویر را می توان ارائه نمود که در سیستم های با نرخ ارسال بسیار پایین مانند خطوط تلفن می توانند مورد استفاده قرار گیرند. چون این روشها عمدتاً در تلفن تصویری بکار می روند هدف اصلی در آنها کد نمودن تصاویر صورت می باشد. بنابراین با در نظر داشتن این پیش فرض برای کد نمودن این تصاویر لازم است محل ویژگی های مهم صورت در یک فریم مشخص شود و تغییرات آنها در فریم های دیگر تعیین گردد. این روش شامل دو مرحله آنالیز و سنتز می باشد. در مرحله آنالیز محل ویژگی های صورت تعیین می گردد و تغییرات آن هم نسبت به فریم های قبلی با استفاده از یک کتاب کد ویژگی ها کد می شود. در مرحله سنتز تصویر با استفاده از کدهای دریافتی و کتاب کد ویژگی های مشابه فرستنده بازسازی می گردد.

در تحقیقات انجام شده از روش "برش و چسباندن"<sup>۱</sup> برای سیستم کدکننده بر مبنای مدل<sup>۲</sup> استفاده شده است. این روش از بلوک های تصاویر چشمها و دهان و یک تصویر کامل صورت استفاده می نماید. مسئله اصلی پیدا کردن محل چشمها و دهان در صورت می باشد تا از این طریق بتوان زیر تصویری از دهان و چشمها در کتاب کد را که بهترین تطبیق را با اعضای مزبور در فریم فعلی دارد برای هر فریم در یک رشته تصاویر انتخاب نمود. به همین جهت روشی برای تعیین محل چشمها ارائه گردیده است که حتی در افراد دارای عینک و یا ریش نیز کاربرد دارد و سپس با استفاده از محل چشمها و روابط اجزاء صورت محل دهان و اندازه بلوک های استفاده شده برای برش زیر تصاویر چشمها و دهان تعیین گردیده است.

**واژه های کلیدی:** کد نمودن بر مبنای مدل، روش برش و چسباندن، کتاب کد ویژگی ها<sup>۳</sup>، گرادیان تصویر

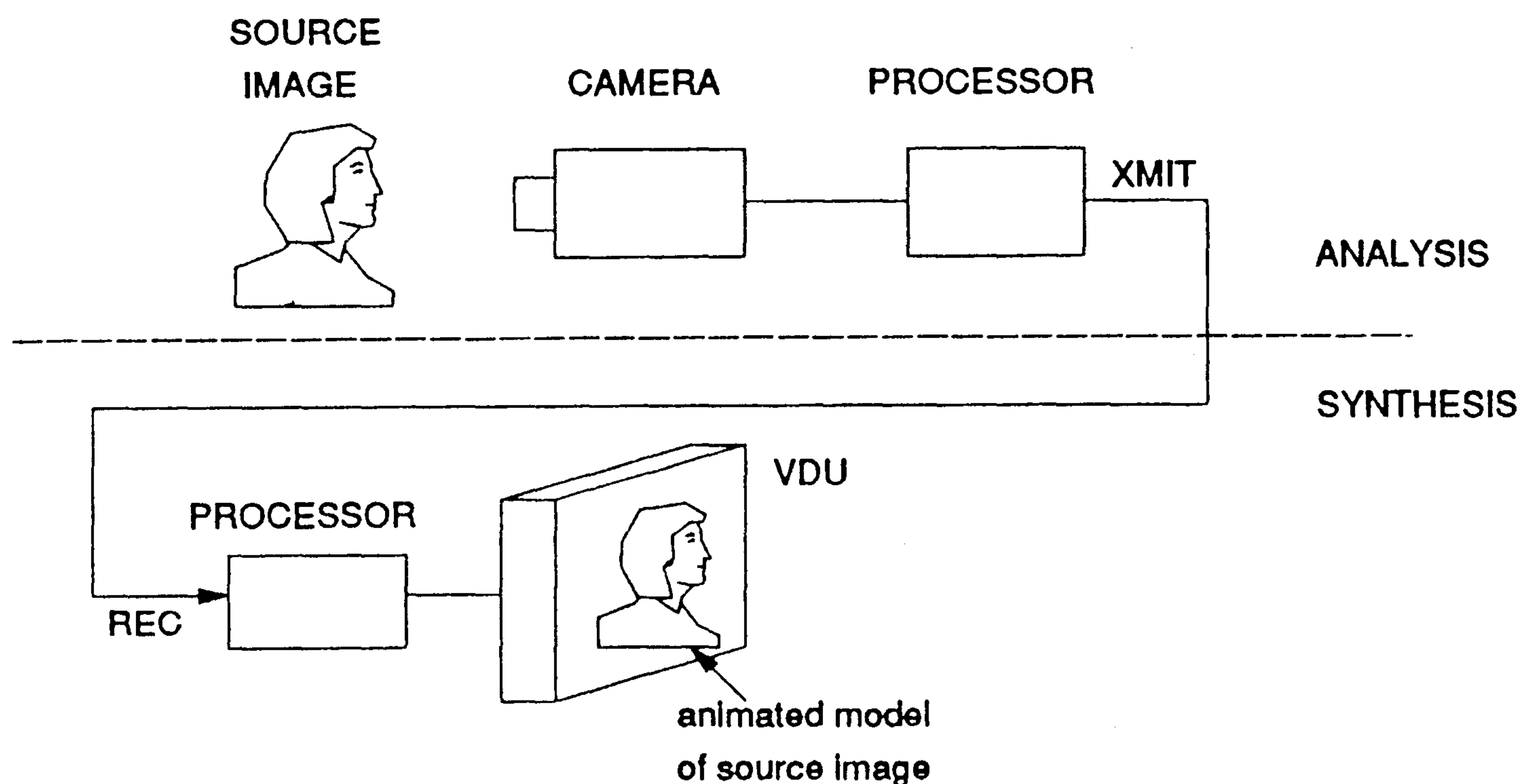
## مقدمه

را تغییر می دهد در هر حال شکل ظاهری سیمای او تغییری نمی کند مگر آنکه جای خود را با شخص دیگری عوض نماید. در صورتی که چنین حالتی رخ ندهد و فرض کنیم که شخص شیء جدیدی را در مقابل دوربین قرار نمی دهد پس از چند فریم تنها لازم است که اطلاعات مربوط به حرکت ارسال شوند. شکل (۱) بلوک دیاگرام یک تلفن تصویری با استفاده از کد کننده بر مبنای مدل را نشان می دهد.

چون این روش عمدتاً در تلفن تصویری بکار می رود هدف اصلی در این حالت کد نمودن تصاویر صورت می باشد. با در نظر

روش های کلاسیک کد نمودن تصاویر پیش فرض های زیادی را در مورد صحنه ای که باید کد شود در نظر نمی گیرند و فقط پیش فرض های کمی در مورد صحنه و خصوصیات سطح اشیاء موجود در آن بطور ضمنی در خصوصیات آماری تصاویر نهفته می باشد. در رشته تصاویر متحرک قابلیت بیشتری برای بکارگیری پیش فرض هایی در مورد تصویر موجود است.

برای مثال یک تصویر در تلفن تصویری شامل سروشانه یک شخص می باشد. در خلال گفتگوی تلفنی شخص دهان خود را حرکت می دهد، سر خود را تکان می دهد و یا حالت های چهره خود



شکل ۱: بلوک دیاگرام تلفن تصویری با استفاده از کدکننده بر مبنای مدل

مشابه خود در تصویر صورت در فریم فعلی دارند انتخاب کرده و به جای تصاویر چشم و دهان در فریم فعلی بکار برده شوند.

### تعیین محل چشم

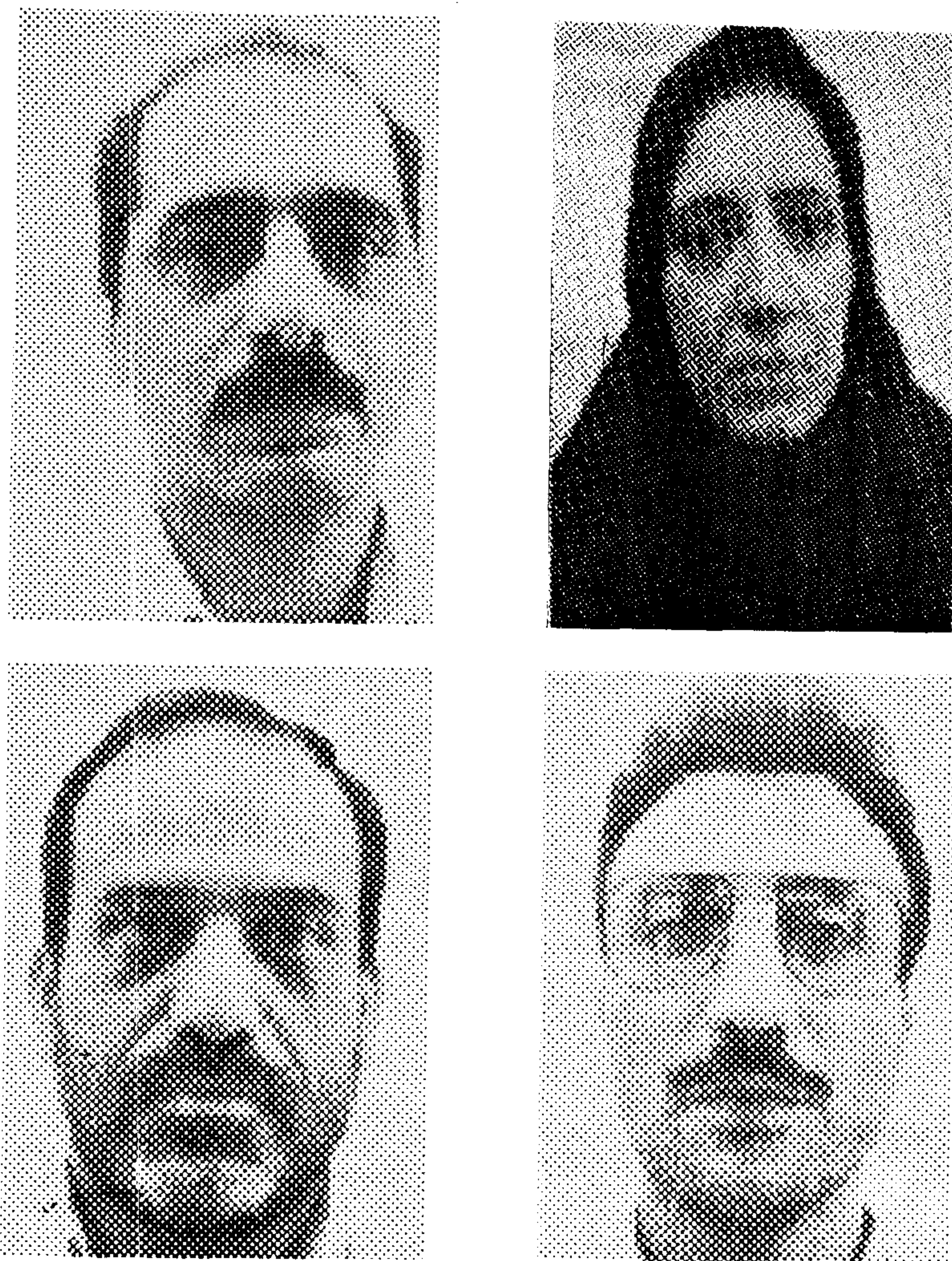
روش‌های متعددی برای تعیین محل ویژگی‌های صورت ارائه گردیده است که هر کدام نقایص و محدودیت‌هایی را دارا می‌باشند. روش ناگاو [۱] در مورد اشخاص دارای ریش و عینک قابل استفاده نیست. در بعضی روش‌ها [۲،۳] تجهیزات اضافی نظیر منبع نور مادون قرمز و دوربین مادون قرمز برای تعیین محل چشم‌ها لازم می‌باشد روشی که در اینجا برای تعیین محل چشم‌ها بکار می‌رود می‌تواند محل چشم‌ها را بدون هیچ محدودیتی آشکار نماید [۴].

مهمترین مشخصه چشم در تصاویر سیاه و سفید تضاد رنگ بین عنبیه و سفیدی چشم می‌باشد. وجود منطقه تیره دایره‌ای شکل مانند عنبیه در پس زمینه سفید در گرادیان تصویر باعث ایجاد گرادیان‌هایی به سمت خارج می‌گردد. به همین جهت اگر خطوطی در جهت عکس گرادیان‌ها رسم گردد این خطوط باید در محل عنبیه با یکدیگر برخورد نمایند. پس از بدست آوردن گرادیان نقاط تصویر [۵]، در صورتی که خطوطی در جهت عکس گرادیان در نقاط تصویر رسم گردد چون این خطوط در نواحی اطراف چشم در نزدیکی عنبیه با یکدیگر تلاقی دارند، برای

گرفتن این پیش‌فرض برای کد نمودن این گونه تصاویر لازم است ویژگی‌های مهم صورت (مانند چشم‌ها و دهان) در یک فریم بدست آیند و تغییرات آنها نسبت به فریم‌های قبل تعیین گردند. بنابراین این روش از دو مرحله آنالیز و سنتز تشکیل گردیده است. در مرحله آنالیز محل ویژگی‌های مهم صورت تعیین می‌گردد و با استفاده از یک کتاب کد تغییرات آنها نسبت به فریم‌های قبل کد می‌شود. در مرحله سنتز، تصویر را می‌توان با استفاده از کدهای دریافتی و کتاب کدی، مشابه آنچه در فرستنده وجود دارد، بازسازی نمود. روش گفته شده بر مبنای در نظر گرفتن مدل دوبعدی برای صورت می‌باشد.

در تجربیات انجام شده، از روش کد کردن بر مبنای مدل با استفاده از کتاب کد ویژگی‌ها برای شبیه‌سازی یک سیستم کد کننده بر مبنای مدل استفاده گردیده است. مسئله اصلی که در این روش مطرح می‌باشد تعیین محل ویژگی‌های صورت (چشم و دهان) در تصاویر تلفن تصویری است. پس از بکارگیری الگوریتمی برای تعیین محل چشم‌ها در تصاویر صورت، تعیین محل دهان و ابعاد استفاده شده برای پنجره‌هایی که جهت برش محل دهان و چشم‌ها بکار می‌روند (بر اساس رابطه اجزاء صورت) انجام می‌گیرد. بعد از تعیین محل تصاویر چشم و دهان باید بوسیله آنها یک کتاب کد ساخته شود و در مرحله جایگزینی تصاویری از کتاب کد که بیشترین انطباق را با تصاویر اجزاء

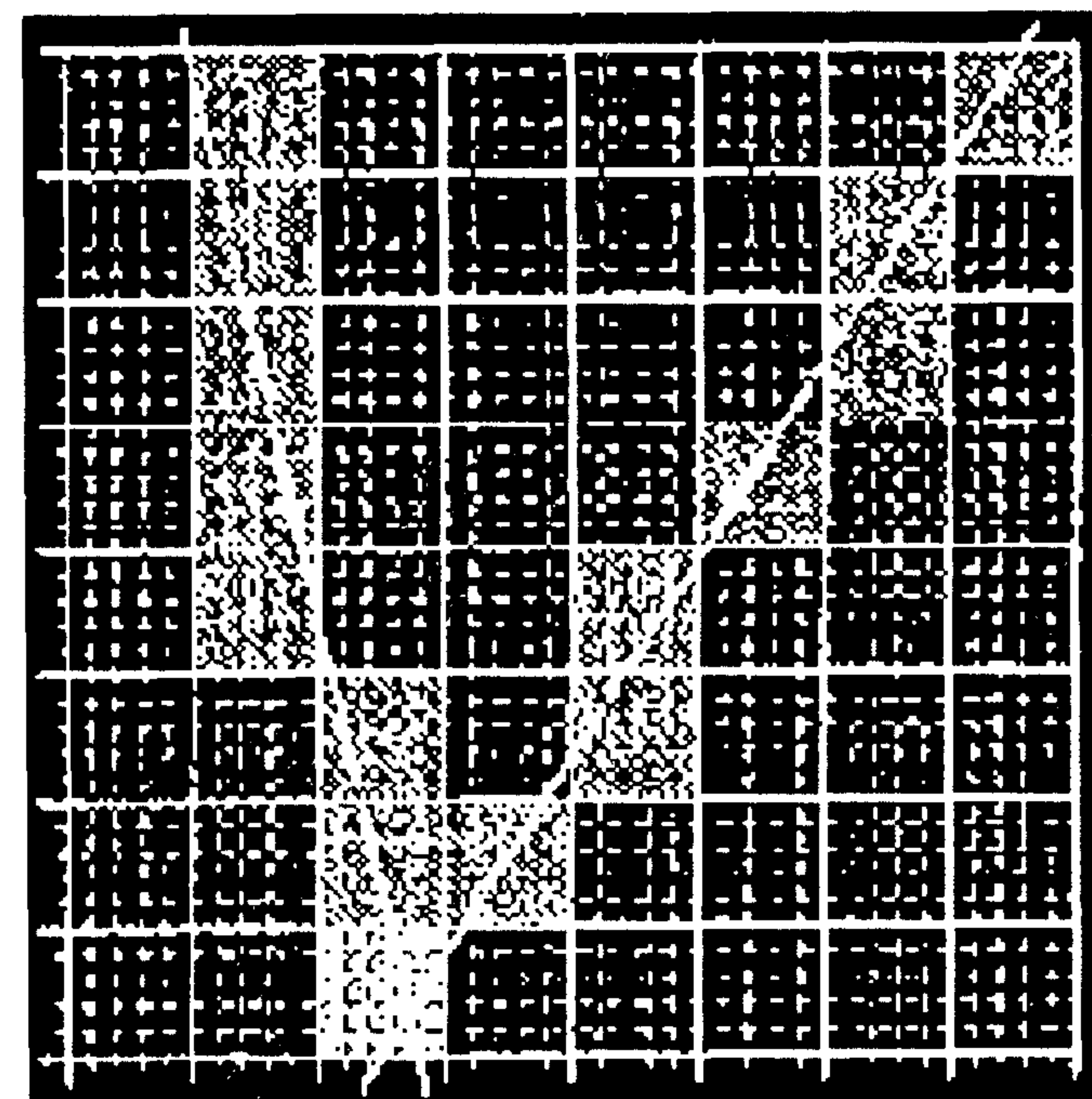
انتخاب اندازه حدود ۵ پیکسل برای ابعاد انباره ضمن آنکه دقت کافی در تعیین محل چشم‌ها فراهم می‌آید، از زمان محاسبه نیز با کارایی بالا استفاده می‌شود. عامل دیگری که کارایی محاسبات را افزایش می‌دهد رسم خطوط در نقاطی است که گرادیان آنها از حد معینی بالاتر می‌باشد. (مثلاً بالاتر از مقدار مؤثر<sup>۲</sup> دامنه گرادیان‌های موجود در شکل) این عمل ضمن آنکه زمان محاسبه را کاهش می‌دهد بخودی خود نقاط اطراف چشم که دارای گرادیان بالاتری نسبت به باقی نقاط تصویر می‌باشند را از دیگر نقاط صورت جدا می‌سازد. نکته دیگری که باید در نظر داشت طول خطوط رسم شده در جهت عکس گرادیان نقاط است برای رسیدن به نتیجه بهتر لازم است که طول این خطوط در حدود ابعادی باشد که برای چشم تصور می‌گردد. در این شبیه‌سازی از تصاویری برای آزمایش استفاده گردیده است که در شکل (۳) نشان داده شده‌اند.



شکل ۳: تصاویر استفاده شده در شبیه‌سازی.

پس از محاسبه گرادیان تصاویر موجود در شکل (۳) مقدار مؤثر گرادیان در این تصاویر محاسبه و نقاطی که دامنه گرادیان آنها بالاتر از مقدار مؤثر بوده و در بدست آوردن محل چشم‌ها بکار رفته‌اند در شکل (۴) نشان داده شده‌اند. سپس از نقاط بدست آمده خطوطی در جهت عکس گرادیان رسم و

مشاهده این اثر یک بردار دو بعدی از انباره‌ها<sup>۱</sup> که در ابتدا به آنها مقدار صفر نسبت داده شده است برای ذخیره نمودن محل عبور این خطوط انتخاب می‌شود. هر گاه خطی در جهت عکس گرادیان یک نقطه رسم شود، این خط از چندین انباره عبور می‌نماید. محتوای هر انباره به ازای یک خط که از آن عبور می‌نماید به میزان یک واحد افزایش داده می‌شود. این فرایند برای دو خط گرادیان در شکل (۲) نشان داده شده است. باید توجه شود که انباره‌هایی که در محل تقاطع خطوط قرار دارند به مقدار بیشتری افزایش می‌یابند که در این شکل بصورت سفیدتر نشان داده شده‌اند. این امر مبنای الگوریتمی است که در اینجا استفاده شده است. واضح است که رسم خطوط در جهت عکس گرادیان در میدان گرادیان اطراف چشم که به سمت خارج است، اثر افزایش مشابهی در انباره‌هایی که تقریباً در مرکز عنبیه قرار دارند ایجاد می‌نماید.



شکل ۲: نمایش نحوه افزایش انباره‌ها برای دو خط دلخواه که در جهت عکس گرادیان رسم گردیده‌اند. انباره‌های روشن‌تر نشان دهنده تجمع بیشتری باشند.

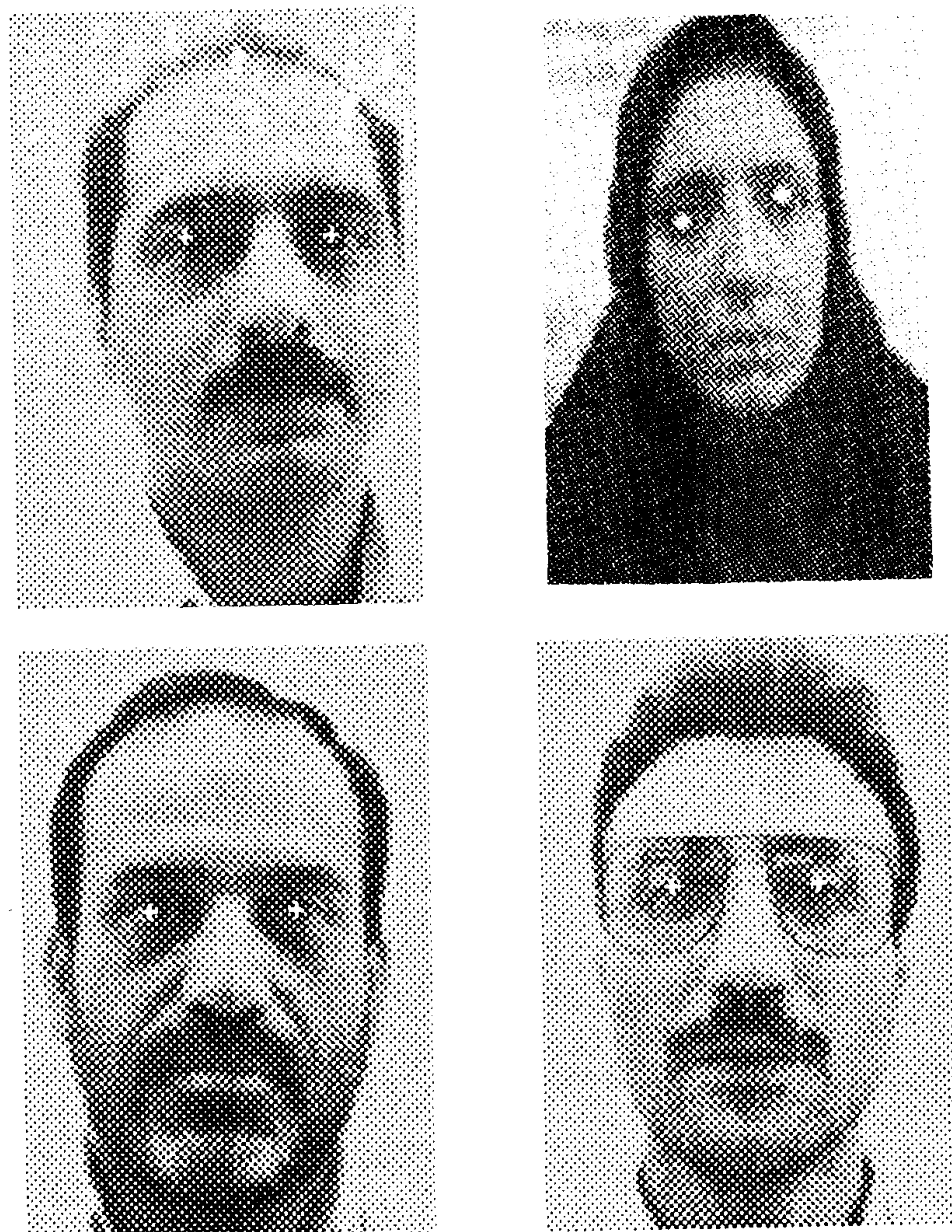
در الگوریتم فوق دو عامل در کیفیت و کارایی این روش بسیار مؤثر می‌باشند. یکی از این عوامل اندازه انباره‌ها برحسب پیکسل می‌باشد. برای انباره‌های بزرگتر زمان محاسبه کمتر می‌گردد. ولی از دقت در تعیین محل چشم نیز کاسته می‌شود. و در صورت انتخاب انباره‌های کوچک ضمن آن که پس از مرحله‌ای این عمل نیز باعث کاهش دقت و زیاد شدن تعداد انباره‌هایی می‌گردد که دارای خطوط عبور کرده زیاد می‌باشند، زمان محاسبه نیز افزایش می‌یابد. در حین آزمایشات مشخص گردیده است که با

پس از رسم خطوط در جهت عکس گرادیان و افزایش محتوای انباره‌ها به ازای هر خطی که از آنها عبور می‌نماید، در حالت ایده‌آل می‌توان انتظار داشت که دو انباره‌ای که متناظر با دو چشم می‌باشند در تصویر بیشترین مقدار را داشته باشد.

اما به علت نویز موجود در زمینه تصویر و حالت‌های مربوط به موهای سر و صورت و طرح‌های لباس ممکن است بیش از دو انباره مقادیر نسبتاً زیادی را ذخیره نمایند. برای انتخاب از میان این انباره‌ها دو پیش فرض در نظر گرفته می‌شود. نخست آنکه محتوای انباره‌های مربوط به دو چشم باید نسبتاً نزدیک به یکدیگر باشد و دیگر آنکه با توجه به اینکه صورت شخص در تصاویر معمولاً در حالت عمودی قرار گرفته است. بنابراین تفاوت بین مؤلفه‌های  $y$  انباره‌های مربوط به چشم‌ها باید اندک باشد در صورتی که محتوای انباره موجود در محل  $(i,j)$  را با  $b_{ij}$  نشان دهیم شرایط فوق را می‌توان با کمینه نمودن تابع زیر ارضا نمود:

$$J = (b_{ij} - b'_{i'j'})^2 + \lambda (j - j')^2$$

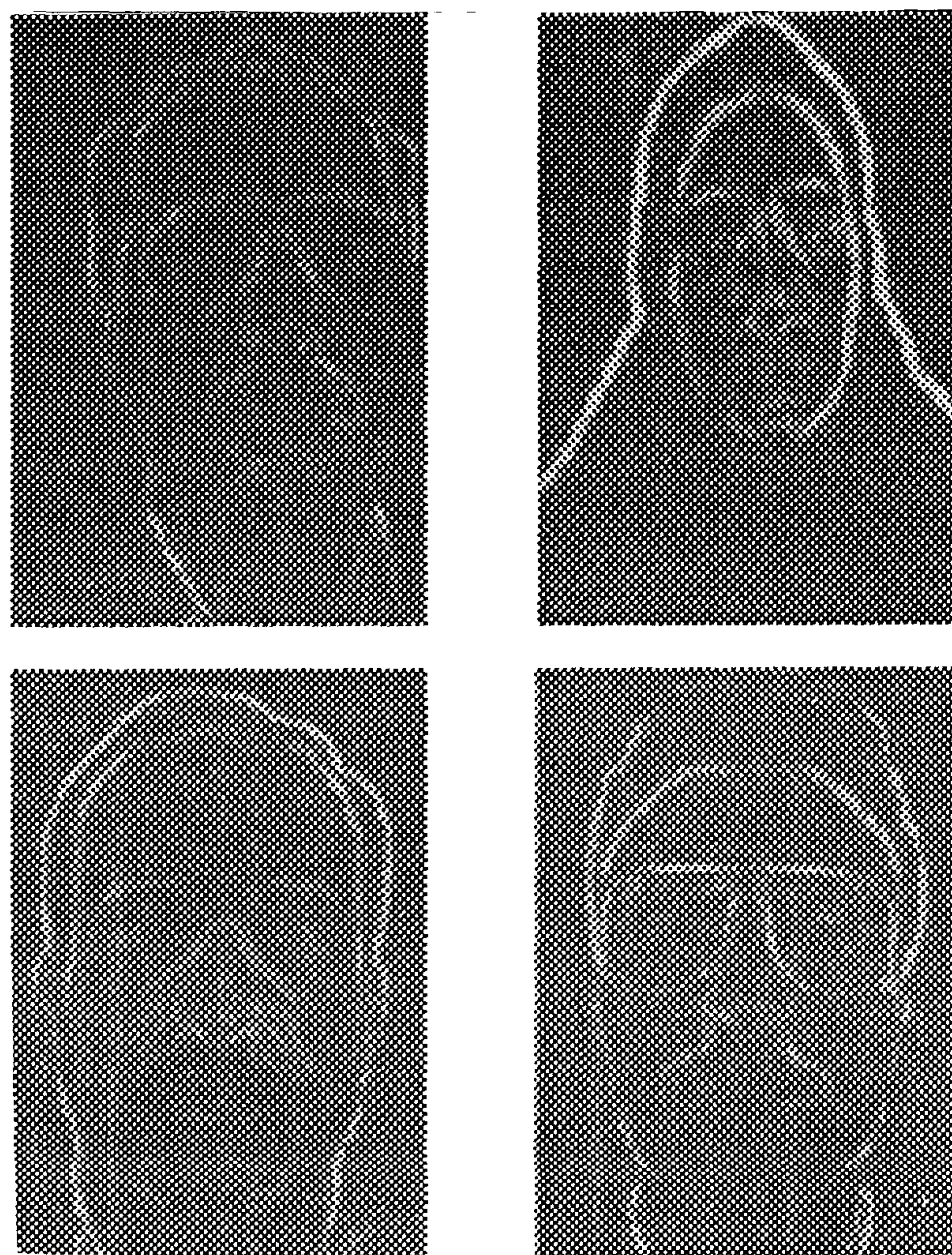
که در این تابع  $(i',j')$ ، نشان دهنده محل دو انباره و  $\lambda$  نشان دهنده وزنی است که به یکی از شرایط فوق نسبت به دیگری داده می‌شود. پس از بکار بردن شرط فوق محل چشم‌ها با دقت نسبتاً خوبی بدست می‌آید نتیجه این شبیه‌سازی در شکل (۶) نشان داده شده است.



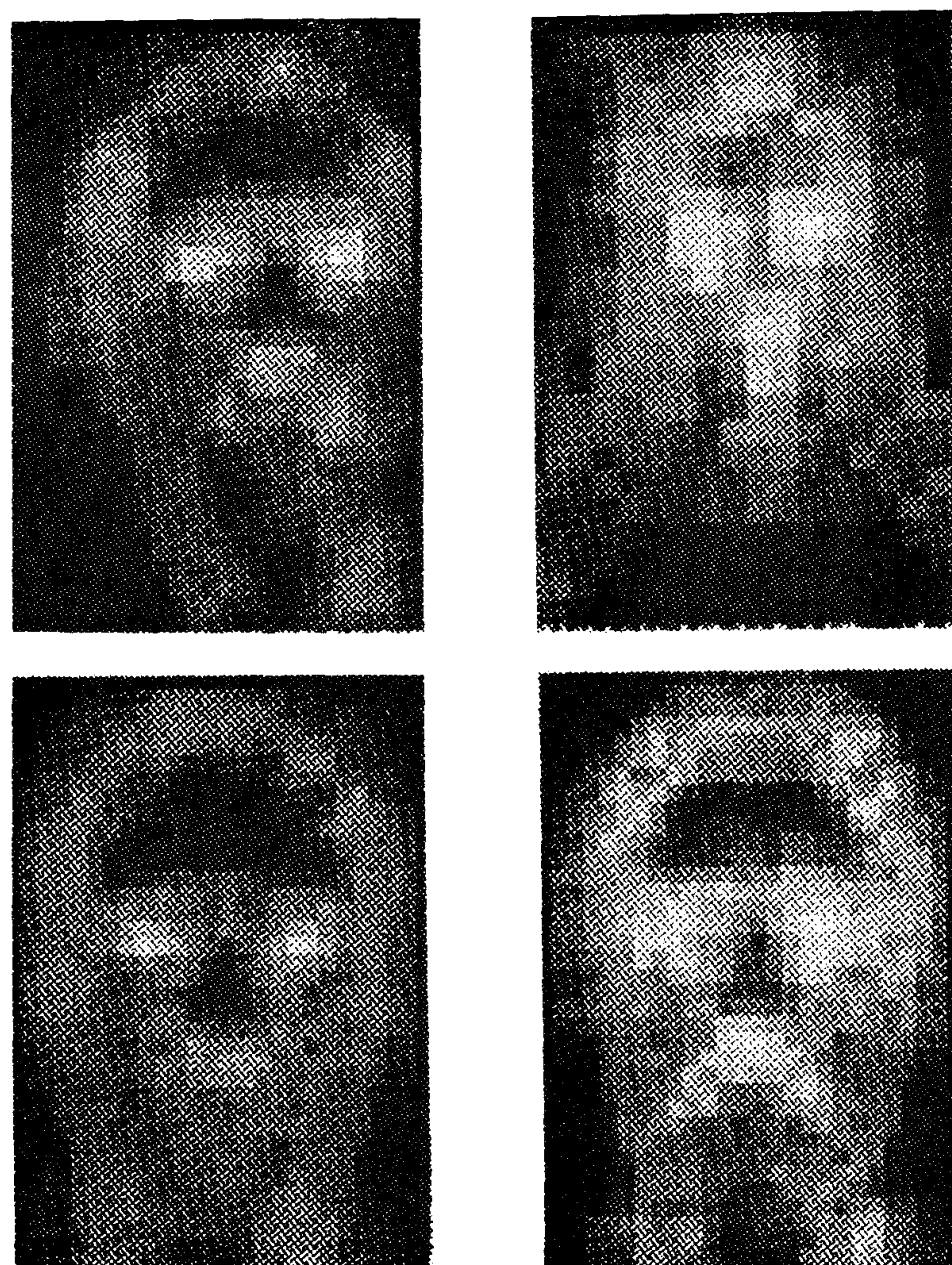
شکل ۶: محل چشم‌ها که بوسیله برنامه پردازش کننده بدست آمده

است

محتوای انباره‌های مربوط به این تصاویر در شکل (۵) نشان داده شده است.



شکل ۴: گرادیان تصاویر استفاده شده در شبیه‌سازی.



شکل ۵: انباره‌های مربوط به تصاویر موجود در شکل (۳).

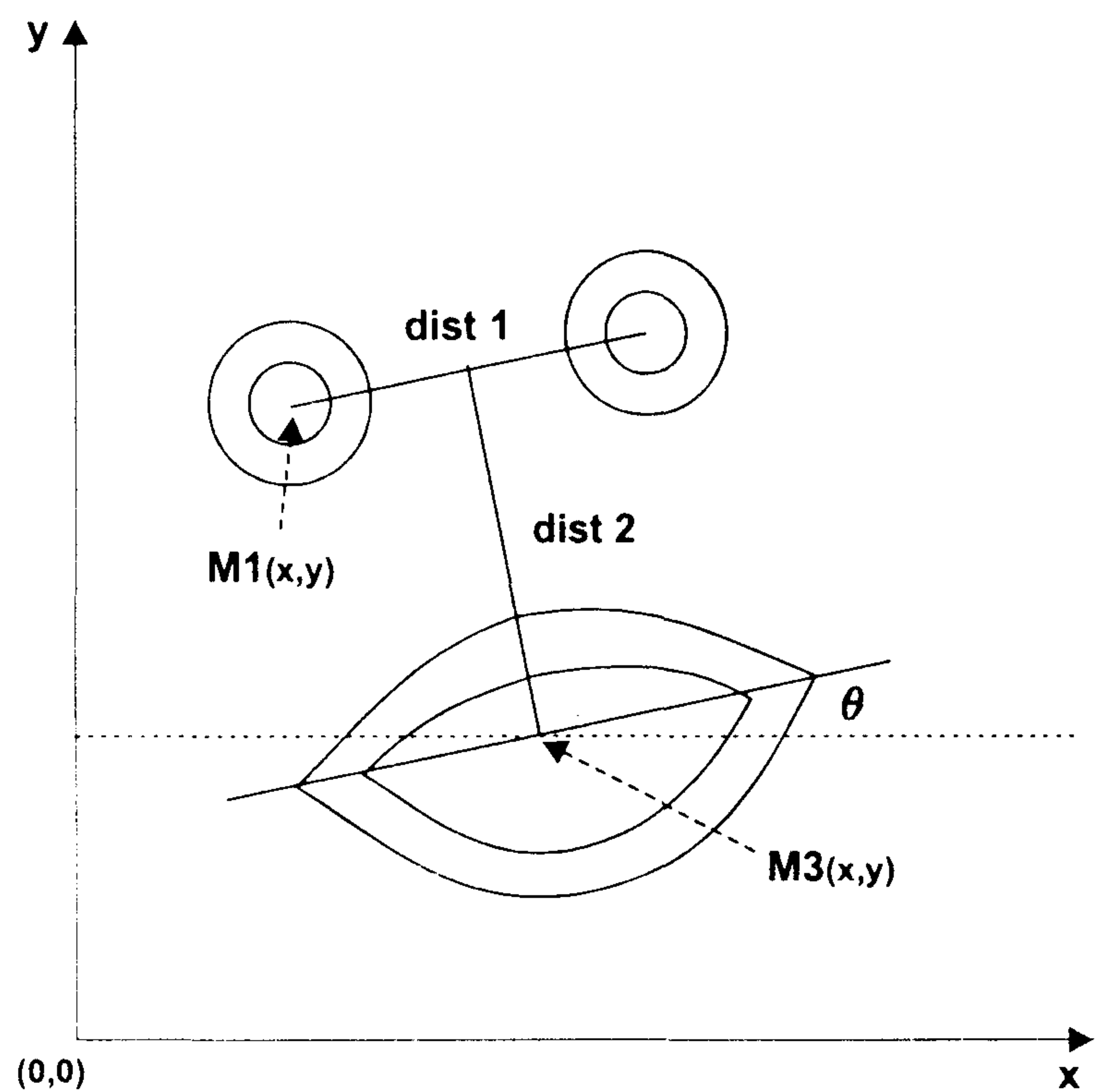
## تخمین محل دهان و ابعاد ویژگی‌های صورت

با مشاهده تصاویر صورت می‌توان روابطی میان اجزاء مختلف صورت مشاهده نمود. برای مثال حداقل فاصله بین مرکز دهان و خط واصل بین مرکز دو چشم تقریباً برابر با فاصله بین دو چشم می‌باشد (شکل ۷) [۶].

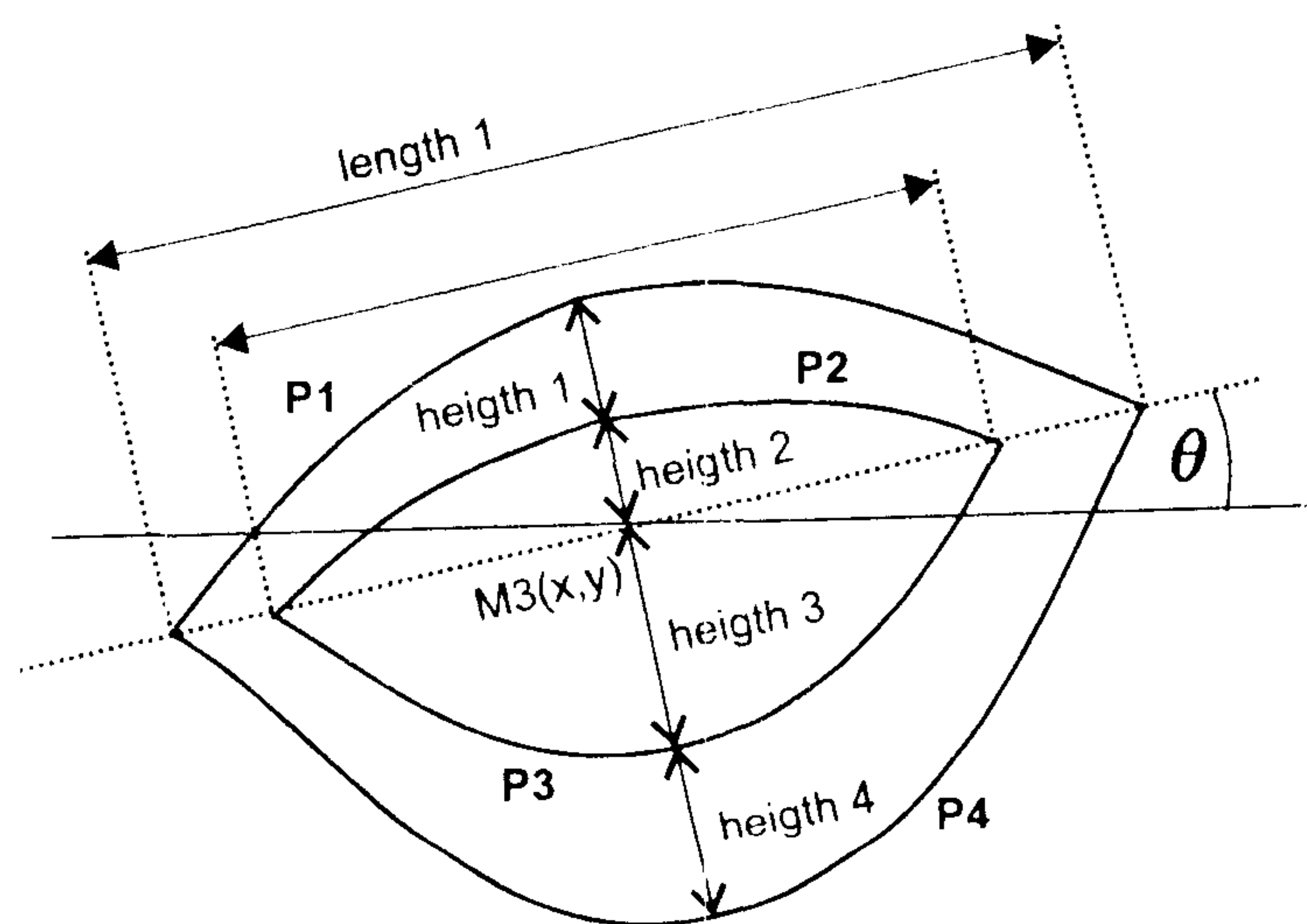
با استفاده از این رابطه می‌توان موقعیت دهان را با داشتن موقعیت چشم‌ها بدست آورد ( $dist2 = 1.1dist1$ ). همچنین طول دهان در حدود فاصله بین دو چشم می‌باشد. بنابراین با توجه به این روابط و داشتن محل چشم، محل دهان در تصاویر مورد آزمایش بدست آمده است (شکل ۸). تخمین ابعاد اجزاء صورت با در نظر گرفتن روابط زیر در شکل (۹) انجام می‌شود:



شکل ۸: محل چشم‌ها و دهان که بوسیله برنامه پردازش کننده بدست آمده است



شکل ۷: مدل صورت انسان.



شکل ۹: مدل دهان انسان.

$$\begin{aligned} length1 &= dist1 & length2 &= 0.7 length1 \\ height1 &= 0.1 length1 & height4 &= 0.16 length1 \\ height2 + height3 &= 0.26 length1 & Max(length1) &= 1.26 dist1 \end{aligned}$$

روابط فوق در تعیین ابعاد پنجره‌ای که برای بلوک مربوط به دهان در روش برش و چسباندن استفاده می‌شود بکار می‌رود و با توجه به این روابط طول بلوک مربوط به دهان  $1.26dist1$  و ارتفاع آن  $0.52dist1 = (0.26 + 0.16 + 0.1)dist1$  می‌باشد. ارتفاع بلوک مربوط به پنجره چشم‌ها در حدود  $0.56dist1$  می‌باشد البته در صورتی که مختصه  $y$  چشم‌ها در تصویر با یکدیگر تفاوت داشته باشد این تفاوت به ارتفاع بلوک مربوط به پنجره چشم‌ها اضافه می‌گردد عرض بلوک پنجره چشم‌ها  $1.7dist1$  در نظر گرفته می‌شود. بلوک‌های مورد نظر برای تصاویر مورد آزمایش در شکل (۱۰) نشان داده شده‌اند.

## پردازش تصاویر تلفن تصویری

در این آزمایش از شخصی درخواست گردید کلمات "یک"، "دو"، "سه" را به آرامی و بطور شمرده بیان نماید. و یک رشته تصاویر متحرک با نرخ ۱۰ فریم در ثانیه و حد تفکیک  $120 \times 160$  پیکسل که دارای ۶۴ سطح خاکستری بود از او در هنگام ادای این کلمات گرفته شد. این رشته تصویر شامل ۳۳ فریم گردید. معیار انتخاب یک تصویر دهان و یا چشم از کتاب کد ویژگی‌ها برای جایگزینی در یک فریم همبستگی بین عضو کتاب کد و عضو

چشم‌ها باید  $۱۶ = ۱۱ + ۵$  پیکسل انتخاب گردد. و پهنای این بلوک  $۳۴ = ۱۷ \times ۲۰$  پیکسل می‌باشد.

مقدار آستانه‌ای برای مجموع قدر مطلق تفاضل سطوح خاکستری در مورد بلوک دهان برابر با ۴۰۰ انتخاب گردید که بطور متوسط برای هر پیکسل تفاوت سطح خاکستری برابر می‌گردد با:

$$\frac{\text{مقدار آستانه‌ای}}{\text{مقدار پیکسل‌ها در بلوک}} = \frac{۴۰۰}{۲۷۵} \cong ۱/۵$$

با انتخاب چنین مقدار آستانه‌ای دهان‌های موجود در رشته تصاویر ضبط شده (۳۳ فریم) بوسیله ۷ بلوک دهان بازسازی گردید. بدلیل آنکه دهان عضو قابل انعطاف‌تری به نسبت چشم‌ها می‌باشد بنابراین برای مقادیر آستانه‌ای مشابه برای دهان و چشم‌ها برای هر پیکسل، تعداد اعضای کتاب کد برای چشم‌ها کمتر از تعداد اعضای کتاب کد برای دهان می‌باشد با انتخاب تفاوت سطح خاکستری برای پیکسل‌های بلوک چشم مشابه با پیکسل‌های بلوک دهان برای بازسازی تصاویر چشم در ۳۳ فریم به ۴ بلوک چشم نیاز می‌باشد. در شکل (۱۱) چهار تصویر اول از تصاویر ضبط شده نشان داده شده است و در شکل (۱۲) تصاویر بازسازی شده متناظر با تصاویر شکل ۱۱ نشان داده شده است. با مشاهده این تصاویر ملاحظه می‌شود که با وجود آنکه وضعیت دهان در دو تصویر اول (دو تصویر سطر اول) اندکی تفاوت دارد ولی در تصاویر بازسازی شده از تصاویر دهان مشابهی برای نمایش آنها استفاده گردیده است و همین امر در مورد تصاویر دهان در دو فریم آخر (دو تصویر سطر دوم) نیز دیده می‌شود.

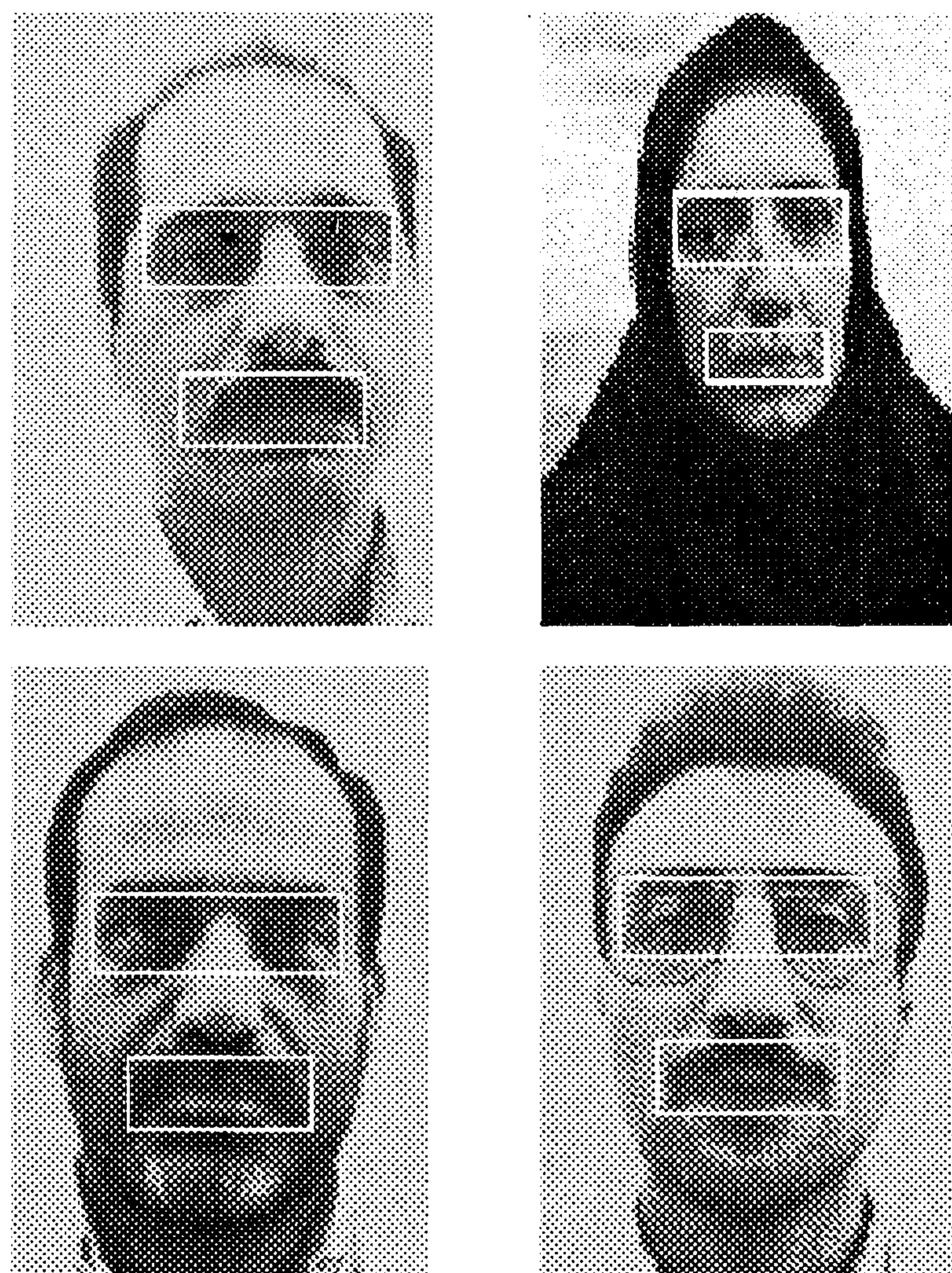


شکل ۱۱: رشته تصویر اصلی.

مربوط در تصویر می‌باشد. همبستگی نیز براساس عکس مجموع قدر مطلق تفاضل سطوح خاکستری بین دو عضو در نظر گرفته شده است.

چون اندازه انباره در تعیین محل چشم‌ها در حدود ۴ پیکسل می‌باشد. بنابراین دقت عملیات تعیین محل چشم‌ها و دهان در حدود ۴ پیکسل است. در نتیجه برای آنکه این تقریب در هنگام قرار دادن بلوک‌های دهان و چشم‌ها بر روی تصویر صورت ایجاد اشکال ننماید، عمل مقایسه هر بلوک با تصویر در داخل پنجره‌ای که ابعاد آن از هر سمت ۴ پیکسل بزرگتر از بلوک مورد نظر می‌باشد انجام می‌گیرد و محلی که همبستگی در آن ماکزیمم می‌باشد به عنوان محل عضو مربوطه در نظر گرفته می‌شود.

در این تصاویر فاصله بین چشم‌ها ۲۰ پیکسل بوده است و با توجه به روابطی که قبلاً ذکر گردید حداقل فاصله خط و اصل بین دو چشم و مرکز دهان برابر با ۲۲ پیکسل می‌گردد. ارتفاع بلوک دهان  $۱۱ \cong ۲۰ \times (۰/۲۶ + ۰/۱۶ + ۰/۱۱)$  پیکسل می‌باشد و پهنای دهان  $۲۵ \cong ۲۰ \times (۱ + ۰/۲۶)$  پیکسل است. ارتفاع بلوک مربوط به چشم‌ها  $۱۱ \cong ۲۰ \times ۰/۵۶$  پیکسل می‌باشد و چون تفاوت بین مؤلفه ۷ چشم‌ها ۵ پیکسل است (چشم‌ها کاملاً در امتداد محور افقی قرار نگرفته‌اند) بنابراین ارتفاع بلوک مربوط به



شکل ۱۰: پنجره‌های مربوط به دهان و چشم‌ها در تصاویر مورد آزمایش.

می‌باشد و شامل  $19200 = 160 \times 120$  پیکسل است و ۶ بلوک دهان شامل  $1650 = 11 \times 25 \times 6$  پیکسل و ۳ بلوک چشم شامل  $1632 = 16 \times 34 \times 3$  پیکسل بکار گرفته شده است بنابراین، در کل بازسازی این تصاویر بوسیله  $22482 = 1650 + 1632 + 19200$  پیکسل انجام گردیده است. در نتیجه نرخ فشرده سازی با استفاده از روش کتاب کد ویژگی‌ها برابر با  $28 \cong \frac{633600}{22482}$  می‌باشد. باید توجه داشت که در کد کردن و ارسال تصاویر مربوط به کتاب کد می‌توان از روش‌های کلاسیک فشرده‌سازی نیز استفاده نمود و به این ترتیب به نرخ‌های فشرده‌سازی بالاتری دست یافت. بنابراین با توجه به اینکه در ارسال تصاویر بر طبق استاندارد H.261 نرخ ارسال 64Kbit/s است. و نسبت فشرده‌سازی در این حالت ۲۸ می‌باشد بنابراین می‌توان انتظار داشت با تلفیق روش‌های فشرده‌سازی کلاسیک با این روش نرخ ارسال را تا حدود ۲Kbit/s پایین آورد. بنابراین اگر چه این روش پیچیدگی محاسباتی را بالا می‌برد ولی به میزان زیادی اطلاعات تصویری را فشرده می‌نماید.

در خاتمه لازم به ذکر است که کیفیت تصاویر بازسازی شده بسیار بالا است به نحوی که کلمات ادا شده را در این تصویر می‌توان «لب خوانی» نمود.



شکل ۱۲: رشته تصاویر بازسازی شده.

### نتیجه گیری

رشته تصاویر اصلی شامل  $633600 = 120 \times 160 \times 33$  پیکسل است و در تصاویر بازسازی شده تصویر یک فریم کامل که خود دارای یک بلوک چشم و یک بلوک دهان

### مراجع

- 1- Welsh, W. J. (1991). *Model-based coding of images*. Ph.D. dissertation, Univ. Essex, Uk.
- 2- Starker, I. et.al. (1990). "A Gaze-Responsive self-Disclosing Display." *Proc. ACM CHI, 90 Human Factors in Computing Conference*, PP. 3-9.
- 3- Hutchinson, T. E. et.al. (1989). "Human-Computer interaction using eye-gaze input." *IEEE Trans. On Sys., Man and Cyber.*, Vol. 19, No. 6, PP. 1527-1534.
- 4- KoThari, R. et.al. (1996). "Detection of eye locations in unconstrained visual images." *IEEE, ICIP96*, PP. 519-522.
- 5- Gonzalez, R. C. and Wintz, P. (1987). *Digital image processing*. Second editions Addison-Wesley publishing company.
- 6-Kober, R. et.al. (1994). "Model-based versus knowledge-guided representation of non-rigid objects : A case study." *IEEE, ICIP94*, PP. 973-977.

### واژه نامه :

- 1 - Cut and Paste Method
- 2 - Model-Based Coding System
- 3 - Features Code Book
- 4 - Bins
- 5 - Root Mean Squared (RMS)

روش برش و چسباندن  
سیستم کد کننده بر مبنای مدل  
کتاب کد ویژگی‌ها  
انباره‌ها  
بالاتر از مقدار موثر